

# Max-linear models on directed acyclic graphs

Nadine Gissibl\* and Claudia Klüppelberg

*Technical University of Munich*

**Abstract:** We consider a new structural equation model in which all random variables can be written as a max-linear function of their parents and independent noise variables. For the corresponding graph we assume that it is a directed acyclic graph. We show that the model is max-linear and detail the relation between the weights of the structural equation model and the max-linear coefficients. We characterize all max-linear models which are generated by this structural equation model. This leads to the presentation of a max-linear structural equation model as the solution of a fixed point equation and to a unique minimal DAG describing the relationships between the variables. The model structure introduces an order between the random variables, which yields certain model reductions, represented by subgraphs of the DAG which we call order DAGs. This results also in a reduced form for the regular conditional distributions compared to previous representations.

Primary 60G70, 60E15, 05C20; secondary 05C75.

**Keywords and phrases:** directed acyclic graph, graphical model, max-linear model, regular conditional distribution, structural equation model.

## 1. Introduction

We define a *graphical model* as a pair  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$ , where the joint probability distribution  $\mathcal{L}(\mathbf{X})$  of the random vector  $\mathbf{X} = (X_1, \dots, X_d)$  is Markov relative to the directed acyclic graph (DAG)  $\mathcal{D}$  (cf. [6], Chapter 3.2). In the literature such models are also called *Bayesian networks* (cf. [5]).

Structural equation models (SEMs) offer a possibility to construct such graphical models (cf. [8, 11]). More precisely, for (measurable) functions  $f_i$  and sets  $\text{pa}(i) \subseteq \{1, \dots, d\} \setminus \{i\}$ , the parents of node  $i$ , define

$$X_i = f_i(\mathbf{X}_{\text{pa}(i)}, Z_i), \quad i = 1, \dots, d, \quad (1.1)$$

where the noise variables  $Z_1, \dots, Z_d$  are jointly independent. The corresponding graph  $\mathcal{D} = (V, E)$  of the SEM (1.1) with node set  $V$  and edge set  $E$  is obtained by drawing directed edges from each variable  $X_k$  for  $k \in \text{pa}(i)$  to  $X_i$ . We require the resulting graph  $\mathcal{D}$  to be a DAG. Throughout the paper we identify the vector  $\mathbf{X}$  with the set of nodes  $V = \{1, \dots, d\}$  and write  $k \rightarrow i$  if there is an edge from  $k$  to  $i$ . From Theorem 1.4.1 of [8] we know that the joint distribution  $\mathcal{L}(\mathbf{X})$  is uniquely defined by the distribution of  $(Z_1, \dots, Z_d)$  and, denoting by  $\text{nd}(v)$  the non-descendants of node  $i$ ,

$$X_i \perp\!\!\!\perp \mathbf{X}_{\text{nd}(i) \setminus \text{pa}(i)} \mid \mathbf{X}_{\text{pa}(i)} \quad (1.2)$$

for all  $i \in V$ ; i.e., the distribution of  $\mathbf{X}$  is Markov relative to  $\mathcal{D}$ , and  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  is a graphical model. Particular attention has been attracted by linear or Gaussian SEMs resulting in a graphical model (see for example [5, 9, 10]).

Our focus is not on sums, but on maxima. Consequently, we introduce a SEM, which is to the best of our knowledge new, defined as

$$X_i = \bigvee_{k \in \text{pa}(i)} c_k^i X_k \vee c_i^i Z_i, \quad i = 1, \dots, d, \quad (1.3)$$

where  $Z_1, \dots, Z_d$  are independent continuous random variables with  $\text{supp}(Z_i) = \mathbb{R}_+ := (0, \infty)$  and  $c_k^i > 0$  for all  $i = 1, \dots, d$  and  $k \in \text{Pa}(i) := \text{pa}(i) \cup \{i\}$ ; the corresponding graph  $\mathcal{D} = (V, E)$  is again required to

\*Support from the TUM Graduate School's International School of Applied Mathematics (ISAM) at Technische Universität München is gratefully acknowledged.

ADDRESS: Center for Mathematical Sciences, Boltzmannstrasse 3, 85748 Garching, Germany,  
e-mail: [n.gissibl@tum.de](mailto:n.gissibl@tum.de); [cklu@tum.de](mailto:cklu@tum.de); url: <http://www.statistics.tum.de>

be a DAG. We call the resulting graphical model  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  *max-linear graphical model* (ML graphical model).

This model is motivated by applications to risk analysis, where extreme risks play an essential role and may propagate through a network. In such a risk setting we may think of the weights  $c_k^i$  as relative quantities, such that a risk may origin with certain proportions in its different ancestors.

In this paper we investigate distributional and graph properties of a ML graphical model  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$ . We show that  $\mathbf{X}$  can be written as a max-linear function of the noise variables and detail the relation between the weights of the representation (1.3) and the max-linear coefficients, which are determined by a path analysis of the DAG  $\mathcal{D}$ . We also characterize all max-linear models which give rise to a ML graphical model. This leads to the presentation of  $\mathbf{X}$  as the solution of a fixed point equation.

It is a simple but important observation that a large noise variable  $Z_i$  may have a lasting influence on the descendants of node  $i$ . Furthermore, since the max-linear coefficients are defined via maxima of weights over different paths, there exist paths which are relevant and we call them *max-weighted*, since they carry these maxima, and others that can be disposed of without changing the model. We present a DAG  $\mathcal{D}^B$  with minimal number of edges such that  $(\mathcal{D}^B, \mathcal{L}(\mathbf{X}))$  is a ML graphical model.

The size of the noise variables in (1.3) gives rise to a certain order between the components of the vector  $\mathbf{X}$ . In particular, if the same noise variable determines the maximum for different nodes, the component variables realize (up to scaling) the same values. Likewise one scaled component can be strictly smaller or larger than another. Possible equalities or strict inequalities between components of  $\mathbf{X}$  decompose the sample space  $\Omega$  in finitely many subspaces, each of which may give rise to a new DAG. It is a natural consequence that a restriction of the sample space given by a specific order between components leads to a complexity reduction of the model.

We translate this into a new concept of order sets and order DAGs, which leads to an almost sure representation of some node by a minimal number of ancestors of this node. The order DAG corresponding to the restriction of the sample space by some specific order in a subset  $O$  of nodes has itself the node set consisting of  $O$  and its ancestors. As edge set we choose all edges, which are relevant with regard to the specific order and are on a max-weighted path in the original DAG  $\mathcal{D}$ . This concept leads to reduced representations of certain nodes, even to almost sure prediction of unknown nodes.

Our paper is organized as follows. In Section 2 we investigate our new ML graphical model in detail and provide in particular necessary conditions on max-linear models to give rise to a ML graphical model  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$ . We also introduce the notion of a max-weighted path, and study its consequences for  $\mathcal{D}$ .

In Section 3 we characterize those max-linear models which give rise to a ML graphical model. Here we also present the DAG  $\mathcal{D}^B$  with minimal number of edges such that  $(\mathcal{D}^B, \mathcal{L}(\mathbf{X}))$  is a graphical model with unique representation (1.3).

Section 4 is devoted to structural properties of  $\mathbf{X}$  like bounds for some component, information given by lowest and highest max-weighted ancestors. This leads to minimal representations of  $\mathbf{X}$  with respect to subsets of nodes of the DAG.

In Section 5 we use these results extensively, when knowing (observing) parts of the DAG and show that they may reduce the complexity of the model considerably. The concept of order sets and order DAGs introduced in this section is based on specific orders between components of the vector  $\mathbf{X}$  and yields model reductions as well as almost sure prediction of some unknown (non-observed) nodes in the DAG.

Finally, in Section 6 we compute regular conditional distribution functions of parts of the vector  $\mathbf{X}$  conditioned on the observation of some part of the vector making extensive use of this complexity reduction. Throughout we illustrate our findings with examples.

**Notation:** We will use the following notation throughout. For a node  $i \in V$  the sets  $\text{an}(i)$ ,  $\text{pa}(i)$ , and  $\text{de}(i)$  denote the *ancestors*, *parents*, and *descendants* of  $i$  with respect to  $\mathcal{D}$ . Furthermore, we use the notation  $\text{An}(i) := \text{an}(i) \cup \{i\}$ ,  $\text{Pa}(i) := \text{pa}(i) \cup \{i\}$ , and  $\text{De}(i) := \text{de}(i) \cup \{i\}$ . For a set  $U \subseteq V$  of nodes we extend this notation in a natural way by writing  $\text{an}(U) = \bigcup_{i \in U} \text{an}(i)$ ,  $\text{An}(U) := \text{an}(U) \cup U$ , and so on.

We write  $U \subseteq V$  for a non-empty subset  $U$  of nodes. For such a subset we then define  $\mathbf{X}_U = (X_i, i \in U)$ . In general, we consider statements for  $i \in \emptyset$  as invalid. Moreover, for arbitrary (possibly random)  $a_i \geq 0$

we set  $\bigvee_{i \in \emptyset} a_i = 0$  and  $\bigwedge_{i \in \emptyset} a_i = \infty$ .

Occasionally, some DAG  $\mathcal{D} = (V, E)$  is *well-ordered*, which means that the set  $V = \{1, \dots, d\}$  of nodes is linearly ordered in a way compatible with  $\mathcal{D}$  such that  $j \in \text{pa}(i)$  implies  $j < i$ .

## 2. Distributional and graph properties of a max-linear graphical model

Let  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  be a ML graphical model. The random vector  $\mathbf{X}$  can be written as a max-linear function of the noise variables. We explain this first for an example, which will appear again later in the paper.

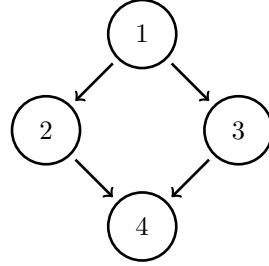
**Example 2.1.** [Every random vector corresponding to a ML graphical model has a max-linear representation]

Consider a ML graphical model  $(\mathcal{D}, \mathcal{L}(X_1, \dots, X_4))$  with

$$\mathcal{D} = (\{1, 2, 3, 4\}, \{(1, 2), (1, 3), (2, 4), (3, 4)\})$$

depicted below. We obtain for the random variables  $X_1, X_2, X_3$  and  $X_4$ :

$$\begin{aligned} X_1 &= c_1^1 Z_1 \\ X_2 &= c_1^2 X_1 \vee c_2^2 Z_2 = c_1^2 c_1^1 Z_1 \vee c_2^2 Z_2 \\ X_3 &= c_1^3 X_1 \vee c_3^3 Z_3 = c_1^3 c_1^1 Z_1 \vee c_3^3 Z_3 \\ X_4 &= c_2^4 X_2 \vee c_3^4 X_3 \vee c_4^4 Z_4 \\ &= c_2^4 (c_1^2 c_1^1 Z_1 \vee c_2^2 Z_2) \vee c_3^4 (c_1^3 c_1^1 Z_1 \vee c_3^3 Z_3) \vee c_4^4 Z_4 \\ &= (c_2^4 c_1^2 c_1^1 \vee c_3^4 c_1^3 c_1^1) Z_1 \vee c_2^4 c_2^2 Z_2 \vee c_3^4 c_3^3 Z_3 \vee c_4^4 Z_4. \end{aligned}$$



The right hand side of this representation of  $\mathbf{X}$  is called *max-linear*; a formal definition is given in (2.4). We summarize the weights of the noise variables in a matrix  $B$ :

$$B = \begin{bmatrix} c_1^1 & c_1^1 c_1^2 & c_1^1 c_1^3 & c_1^1 c_1^2 c_2^4 \vee c_1^1 c_1^3 c_3^4 \\ 0 & c_2^2 & 0 & c_2^2 c_2^4 \\ 0 & 0 & c_3^3 & c_3^3 c_3^4 \\ 0 & 0 & 0 & c_4^4 \end{bmatrix}.$$

Note that  $B$  is upper triangular, since the DAG  $\mathcal{D}$  is well-ordered.  $\square$

We provide a general method to calculate the max-linear representation of  $\mathbf{X}$  by a path analysis. A (*directed*) *path* in  $\mathcal{D}$  from  $j$  to  $i$  is a sequence of distinct nodes in  $V$  such that  $j = k_0 \rightarrow k_1 \rightarrow \dots \rightarrow k_{n-1} \rightarrow k_n = i$  for some  $n \in \mathbb{N}$ . We write such a path

$$p = [j \Rightarrow i] = [j = k_0 \rightarrow k_1 \rightarrow \dots \rightarrow k_{n-1} \rightarrow k_n = i]$$

and denote by  $P_{ji}$  the set of all paths from  $j$  to  $i$ .

**Theorem 2.2.** Let  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  be a ML graphical model. Define for a path  $p = [j = k_0 \rightarrow k_1 \rightarrow \dots \rightarrow k_n = i]$  in  $\mathcal{D}$  the constants

$$d_{ji}^p := c_{k_0}^{k_0} c_{k_0}^{k_1} \dots c_{k_{n-2}}^{k_{n-1}} c_{k_{n-1}}^{k_n} = c_{k_0}^{k_0} \prod_{l=0}^{n-1} c_{k_l}^{k_{l+1}} \quad (2.1)$$

and set

$$b_{ji} = \bigvee_{p \in P_{ji}} d_{ji}^p, \quad b_{ii} = c_i^i \quad \text{for all } j \in \text{an}(i) \text{ and } i = 1, \dots, d \quad (2.2)$$

and all other  $b_{ji} = 0$ . Then

$$X_i = \bigvee_{j \in \text{An}(i)} b_{ji} Z_j, \quad i = 1, \dots, d. \quad (2.3)$$

*Proof.* First note that wlog we can assume that the DAG  $\mathcal{D} = (V, E)$  is well-ordered. We prove the identity (2.3) by induction on the number of nodes of  $\mathcal{D}$ . For  $d = 1$  we have

$$X_1 = c_1^1 Z_1 = b_{11} Z_1,$$

such that both representations are equal provided that  $b_{11} = c_1^1$ . Assume that (2.3) holds for a ML graphical model  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  of dimension  $d$ ; i.e.,

$$X_k = \bigvee_{j \in \text{pa}(k)} c_j^k X_j \vee c_k^k Z_k = \bigvee_{j \in \text{An}(k)} b_{jk} Z_j = \bigvee_{j \in \text{an}(k)} \bigvee_{p \in P_{jk}} d_{jk}^p Z_j \vee c_k^k Z_k, \quad k = 1, \dots, d.$$

Now consider a well-ordered DAG  $\mathcal{D}$  corresponding to a ML graphical model with  $d + 1$  nodes, and note that for  $i = 1, \dots, d$  we have  $(d + 1) \notin \text{pa}(i)$ . In order to verify (2.3) for the nodes  $i = 1, \dots, d$ , it is therefore sufficient to consider the subgraph  $\mathcal{D}[\{1, \dots, d\}] = (\{1, \dots, d\}, E \cap (\{1, \dots, d\} \times \{1, \dots, d\}))$ . Due to the induction hypothesis, (2.3) holds for  $\mathcal{D}[\{1, \dots, d\}]$  and therefore also for  $\mathcal{D}$ . So we can use this hypothesis and (A.1) to obtain

$$\begin{aligned} X_{d+1} &= \bigvee_{k \in \text{pa}(d+1)} c_k^{d+1} X_k \vee c_{d+1}^{d+1} Z_{d+1} \\ &= \bigvee_{k \in \text{pa}(d+1)} \bigvee_{j \in \text{an}(k)} \bigvee_{p \in P_{jk}} c_k^{d+1} d_{jk}^p Z_j \vee \bigvee_{k \in \text{pa}(d+1)} c_k^{d+1} c_k^k Z_k \vee c_{d+1}^{d+1} Z_{d+1} \\ &= \bigvee_{j \in \text{an}(d+1)} \left( \bigvee_{k \in \text{de}(j) \cap \text{pa}(d+1)} \bigvee_{p \in P_{jk}} c_k^{d+1} d_{jk}^p \vee \bigvee_{k \in \text{pa}(d+1) \cap \{j\}} c_k^{d+1} c_k^k \right) Z_j \vee c_{d+1}^{d+1} Z_{d+1}. \end{aligned}$$

For some path  $p$  from  $j$  to  $d + 1$  of the form  $[j \Rightarrow k \rightarrow d + 1]$  for some  $k \in \text{de}(j) \cap \text{pa}(d + 1)$ , we have  $d_{j,d+1}^p = d_{jk}^p c_k^{d+1}$ , and for some edge  $[j \rightarrow d + 1]$  we must have  $d_{j,d+1}^{[j \rightarrow d+1]} = c_j^j c_j^{d+1}$ . Observe that all paths from some  $j \in \text{an}(d + 1)$  to  $d + 1$  satisfy one of these two presentations. This yields

$$X_{d+1} = \bigvee_{j \in \text{an}(d+1)} \bigvee_{p \in P_{j,d+1}} d_{j,d+1}^p Z_j \vee c_{d+1}^{d+1} Z_{d+1} = \bigvee_{j \in \text{An}(d+1)} b_{j,d+1} Z_j,$$

where we have set  $b_{d+1,d+1} = c_{d+1}^{d+1}$  and  $b_{j,d+1} = \bigvee_{p \in P_{j,d+1}} d_{j,d+1}^p$  for all  $j \in \text{an}(d + 1)$ .  $\square$

In the following we study the random vector  $\mathbf{X}$  as a max-linear model with specific properties induced by the graph  $\mathcal{D}$ . We call a random vector  $\mathbf{X}$  *max-linear* if there exist independent continuous random variables  $Z_1, \dots, Z_d$  with  $\text{supp}(Z_i) = \mathbb{R}_+ := (0, \infty)$  and a matrix  $B = (b_{ij})_{d \times d}$ , which we call *max-linear coefficient matrix* (ML coefficient matrix), with non-negative entries such that

$$X_i = \bigvee_{j=1}^d b_{ji} Z_j, \quad i = 1, \dots, d, \quad (2.4)$$

(for background on max-linear models in the context of extreme value theory see for example [2], Chapter 6).

Max-linear models can be linked in a natural way to ML graphical models.

**Proposition 2.3.** *Let  $\mathbf{X}$  be generated by a SEM (1.1) whose corresponding graph is a DAG  $\mathcal{D} = (V, E)$ , and assume that  $\mathbf{X}$  is max-linear. Then the components of  $\mathbf{X}$  satisfy (1.3) and  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  is a ML graphical model.*

*Proof.* For  $i \in V$  we substitute the parents of  $X_i$  by their corresponding representation (1.1) and proceed recursively. After finitely many steps we obtain  $X_i = g_i(\mathbf{Z}_{\text{An}(i)})$  for appropriate functions  $g_i$ . Comparing  $X_i = g_i(\mathbf{Z}_{\text{An}(i)})$  and the max-linear representation  $X_i = \bigvee_{j=1}^d b_{ji} Z_j$  yields that we may set  $b_{ji} = 0$  if  $j \notin \text{An}(i)$  and, hence,  $X_i = \bigvee_{j \in \text{An}(i)} b_{ji} Z_j$ . We therefore have the two representations

$$X_i = \bigvee_{j \in \text{An}(i)} b_{ji} Z_j = f_i(\mathbf{X}_{\text{pa}(i)}, Z_i).$$

Since for every  $k \in \text{pa}(i)$  we have  $X_k = b_{j(k)k} Z_{j(k)}$  for some  $j(k) \in \text{An}(k)$ , we obtain

$$X_i = \bigvee_{j \in \text{An}(i)} b_{ji} Z_j = f_i(\mathbf{X}_{\text{pa}(i)}, Z_i) = h_i(\mathbf{Z}_{\{j(k): k \in \text{pa}(i)\}}, Z_i)$$

for appropriate functions  $h_i$ , where the change from  $f_i$  to  $h_i$  takes care of the coefficients  $b_{j(k)k}$  for  $k \in \text{pa}(i)$ . Thus we can write the max-linear representation as

$$\begin{aligned} X_i &= \bigvee_{k \in \text{pa}(i)} b_{j(k)i} Z_{j(k)} \vee b_{ii} Z_i \\ &= \bigvee_{k \in \text{pa}(i)} \frac{b_{j(k)i}}{b_{j(k)k}} b_{j(k)k} Z_{j(k)} \vee b_{ii} Z_i \\ &= \bigvee_{k \in \text{pa}(i)} \frac{b_{j(k)i}}{b_{j(k)k}} X_k \vee b_{ii} Z_i, \end{aligned}$$

which is (1.3) with  $c_k^i = \frac{b_{j(k)i}}{b_{j(k)k}}$  and  $c_i^i = b_{ii}$ .  $\square$

For what follows we need some graph theoretical notions. The *reachability matrix*  $R = (r_{ij})_{d \times d}$  of a DAG  $\mathcal{D}$  is the matrix

$$r_{ji} = \begin{cases} 1, & \text{if there is a path from } j \text{ to } i, \\ 0, & \text{otherwise.} \end{cases}$$

If  $r_{ji} = 1$ , we say  $i$  is *reachable* from  $j$ , and we set for practical reasons  $r_{ii} = 1$  for  $i = 1, \dots, d$ . By Theorem 2.2 we have for the ML coefficient matrix  $B = (b_{ij})_{d \times d}$  corresponding to a ML graphical model that  $b_{ji} \neq 0$  if and only if  $i$  is reachable from  $j$  or, equivalently, if and only if  $j \in \text{An}(i)$ .

**Remark 2.4.** Let  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  be a ML graphical model.

- (i) Between the ML coefficient matrix  $B$  of  $\mathbf{X}$  and the reachability matrix  $R$  of  $\mathcal{D}$  the following relation holds:

$$R = \text{sgn}(B),$$

where we define equality of matrices componentwise. Hence, the ML coefficient matrix  $B$  is a weighted reachability matrix of the DAG  $\mathcal{D}$ .

- (ii) Moreover, if the DAG  $\mathcal{D}$  is well-ordered, then  $R$  is an upper triangular matrix.  $\square$

In the following we investigate further the relationship between the weights in (1.3) and the ML coefficient matrix  $B = (b_{ij})_{d \times d}$  in (2.3). To this end we define the following matrix operation

$$\odot : \mathbb{R}^{p_1 \times d} \times \mathbb{R}^{d \times p_2} \rightarrow \mathbb{R}^{p_1 \times p_2}, \quad (F, G) \mapsto F \odot G =: H \quad \text{such that} \quad \bigvee_{k=1}^d f_{ik} g_{kj} \mapsto h_{ij}. \quad (2.5)$$

Denote  $F^{\odot 0} = \text{id}_{d \times d}$  and  $F^{\odot n} = F^{\odot(n-1)} \odot F$  for  $n \geq 1$ .

**Theorem 2.5.** Let  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  be a ML graphical model with weights  $c_k^i$  for  $i = 1, \dots, d$  and  $k \in \text{Pa}(i)$  and ML coefficient matrix  $B = (b_{ij})_{d \times d}$ . Define the matrices

$$D = (d_{ij})_{d \times d} := \text{diag}(c_i^i, i = 1, \dots, d), \quad D_0 = (d_{0,ji})_{d \times d} := (c_j^i \mathbf{1}([j \rightarrow i]))_{d \times d},$$

and  $D_1 = (d_{1,ji})_{d \times d} := (c_j^i c_j^j \mathbf{1}([j \rightarrow i]))_{d \times d}$ . Then the matrix  $B$  has representation

$$B = D \quad \text{for } d = 1 \quad \text{and} \quad B = D \vee \bigvee_{k=0}^{d-2} (D_1 \odot D_0^{\odot k}) \quad \text{for } d \geq 2,$$

where  $\vee$  denotes componentwise maxima.

*Proof.* For  $d = 1$  we know from (2.2) that  $b_{11} = c_1^1$ . Hence,  $B = D$ . Now assume that  $d \geq 2$ . We denote by  $P_{ji}^n$  the set of all paths of length  $n \in \mathbb{N}$  (defined by  $n$  edges) from  $j$  to  $i$ . Recall from (2.1) and (2.2) that the coefficients  $b_{ji}$  are defined via maxima over the different paths from  $j$  to  $i$ . Splitting up this operation into maxima over paths of fixed length  $n$  we obtain

$$b_{ji} = \bigvee_{n=1}^{\infty} \bigvee_{p \in P_{ji}^n} d_{ji}^p, \quad j \in \text{an}(i).$$

First, we show that if there exist paths of length  $n$  from  $j$  to  $i$  the  $ji$ -th component of the matrix  $D_1 \odot D_0^{\odot n-1}$  equals the maximal weight of these paths, otherwise it is zero. For an edge  $p = [j \rightarrow i]$ , which is the only path of length  $n = 1$ , we obtain  $d_{ji}^p = c_j^j c_j^i$ . Observe that the  $ji$ -th component of the matrix  $D_1 \odot D_0^{\odot 0} = D_1 \odot \text{id}_{d \times d} = D_1$  is given by  $d_{1,ji} = c_j^j c_j^i \mathbf{1}([j \rightarrow i])$ . So the statement is true for  $n = 1$ . We proceed by induction on  $n \in \mathbb{N}$ , assuming that the statement is true for  $n$ . For a path  $p = [j \rightarrow k_1 \rightarrow \dots \rightarrow k_n \rightarrow i]$  from  $j$  to  $i$  of length  $n + 1$  for arbitrary nodes  $k_1, \dots, k_n$  we obtain  $d_{ji}^p = c_j^j c_j^{k_1} \dots c_{k_{n-1}}^{k_n} c_{k_n}^i$  and, therefore,

$$\bigvee_{p \in P_{ji}^{n+1}} d_{ji}^p = \bigvee_{\{k_1, \dots, k_n: [j \rightarrow k_1 \rightarrow \dots \rightarrow k_n \rightarrow i]\}} c_j^j c_j^{k_1} \dots c_{k_{n-1}}^{k_n} c_{k_n}^i.$$

Let  $d_{n,ji}$  and  $d_{n+1,ji}$  be the  $ji$ -th component of  $D_1 \odot D^{\odot(n-1)}$  and  $D_1 \odot D^{\odot n}$ , respectively. Since  $D_1 \odot D_0^{\odot n} = (D_1 \odot D^{\odot(n-1)}) \odot D_0$ , we have  $d_{n+1,ji} = \bigvee_{k=1}^d d_{n,jk} d_{0,ki} = \bigvee_{k=1}^d d_{n,jk} c_k^k \mathbf{1}([k \rightarrow i])$ . Assuming the existence of a path of length  $n$  from  $j$  to  $k$  and an edge  $[k \rightarrow i]$ , we obtain from the induction hypothesis that

$$d_{n,jk} d_{0,ki} = d_{n,jk} c_k^k = \bigvee_{\{k_1, \dots, k_{n-1}: [j \rightarrow k_1 \rightarrow \dots \rightarrow k_{n-1} \rightarrow k \rightarrow i]\}} c_j^j c_j^{k_1} \dots c_{k_{n-1}}^{k_n} c_k^i,$$

otherwise we have  $d_{n,jk} d_{0,ki} = 0$ . In particular, all paths from  $j$  to  $i$  of length  $n + 1$  are of this form for some node  $k$ . Thus, if there exist paths of length  $n + 1$  from  $j$  to  $i$ , the  $ji$ -th component  $d_{n+1,ji} = \bigvee_{k=1}^d d_{n,jk} d_{0,ki}$  of  $D_1 \odot D^{\odot n}$  is indeed equal to the maximal weight of these paths, otherwise it is zero.

Noting that due to the acyclicity a path in a DAG with  $d$  nodes does not contain more than  $d - 1$  edges we obtain that for  $j \in \text{an}(i)$  the  $ji$ -th component of  $\bigvee_{k=0}^{d-2} D_1 \odot D_0^{\odot k}$  equals the maximal weight of all paths from  $j$  to  $i$ , otherwise it is zero. Recall from (2.2) that  $b_{ji} = 0$  if  $j \notin \text{An}(i)$  and  $b_{ii} = c_i^i$ . This is taken care of by the diagonal matrix  $D$ . Thus the ML coefficient matrix  $B$  is given by

$$B = D \vee D_1 \vee (D_1 \odot D_0) \vee (D_1 \odot D_0^{\odot 2}) \vee \dots \vee (D_1 \odot D_0^{\odot(d-2)}).$$

□

The definition of the matrix operation  $\odot$  in (2.5) modifies and extends the definition given in Section 2.1 of [12]. For some ML graphical model  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  with ML coefficient matrix  $B = (b_{ij})_{d \times d}$  summarizing the noise variables of  $\mathbf{X}$  into the row vector  $\mathbf{Z} = (Z_1, \dots, Z_d)$ , the matrix operation reduces to

$$\mathbf{X}^T = \mathbf{Z} \odot B = \left( \bigvee_{j=1}^d b_{ji} Z_j, i = 1, \dots, d \right) = \left( \bigvee_{j \in \text{An}(i)} b_{ji} Z_j, i = 1, \dots, d \right).$$

The coefficients of the max-linear representation (2.3) of  $\mathbf{X}$  are determined as maxima of products along all paths from  $j$  to  $i$  (cf. (2.1) and (2.2)). This suggests the following definition.

**Definition 2.6.** Let  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  be a ML graphical model with ML coefficient matrix  $B = (b_{ij})_{d \times d}$ . For  $j \in \text{an}(i)$  and  $i \in V$  we call a path  $p = [j = k_0 \rightarrow k_1 \rightarrow \dots \rightarrow k_n = i]$  from  $j$  to  $i$  *max-weighted* if  $b_{ji} = c_{k_0}^{k_0} \prod_{l=0}^{n-1} c_{k_l}^{k_{l+1}} = d_{ji}^p$ . Obviously, there may exist several max-weighted paths from  $j$  to  $i$ , and we denote by  $P_{ji}^{\text{mw}}$  the set of all max-weighted paths from  $j$  to  $i$ . □

**Remark 2.7.** Assume the situation of Definition 2.6. Let  $i \in V$  and  $j \in \text{An}(i)$ . Then following statements are immediate consequences from the definition of max-weighted paths and (2.1).

- (i) A path  $p$  from  $j$  to  $i$  is max-weighted if and only if  $b_{ji} = d_{ji}^p$ .
- (ii) A path  $p$  from  $j$  to  $i$  is not max-weighted if and only if  $b_{ji} > d_{ji}^p$ .

- (iii) Let  $p \in P_{ji}^{\text{mw}}$ . Then every sub-path of  $p$  is also max-weighted.
- (iv) Let  $l_1, l_2 \in V$  and  $p \in P_{ji}^{\text{mw}}$  such that  $p$  contains some sub-path  $p_1$  from  $l_1$  to  $l_2$ . Then the path which results from  $p$  by replacing  $p_1$  by some  $p_2 \in P_{l_1 l_2}^{\text{mw}}$  is also max-weighted.  $\square$

Remark 2.7(i) indicates that for  $i \in V$  the max-linear representation  $X_i = \bigvee_{j \in \text{An}(i)} b_{ji} Z_j$  only depends on the weight  $b_{ii}$  and the weights along max-weighted paths from ancestors of  $i$  to  $i$ .

We discuss Example 2.1 in the light of Definition 2.6 and this observation.

**Example 2.8.** [Continuation of Example 2.1: max-weighted paths]

Assume that  $b_{14} = \frac{b_{12}b_{24}}{b_{22}} > \frac{b_{13}b_{34}}{b_{33}}$ , which is equivalent to  $c_1^1 c_1^2 c_2^4 > c_1^1 c_1^3 c_3^4$ . Thus we have  $P_{14}^{\text{mw}} = \{[1 \rightarrow 2 \rightarrow 4]\}$ ; i.e., the only max-weighted path from 1 to 4 contains node 2, and there is no max-weighted path which contains node 3. Observe from this that the DAG  $(\{1, 2, 3, 4\}, \{(1, 2), (2, 4), (3, 4)\})$  completely determines the max-linear representation of  $X_4 = b_{14}Z_1 \vee b_{24}Z_2 \vee b_{34}Z_3 \vee b_{44}Z_4$ .  $\square$

This important observation motivates the following.

**Definition 2.9.** Let  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  be a ML graphical model, and let  $A \subseteq V$ .

- (a) For some  $i \in V$  we call an ancestor  $j$  of  $i$  *max-weighted by  $A$*  if there exists a max-weighted path from  $j$  to  $i$  which contains a node of  $A$ . We denote by  $\text{an}_{\text{mw}}^A(i)$  the set of the ancestors of  $i$  which are max-weighted by  $A$ . By  $\text{an}_{\text{nmw}}^A(i)$  we further denote the set of the ancestors of  $i$  which are *not max-weighted* by  $A$ ; i.e.,  $\text{an}_{\text{nmw}}^A(i) = \text{an}(i) \setminus \text{an}_{\text{mw}}^A(i)$ .
- (b) For some  $j \in V$  we call a descendant  $i$  of  $j$  *max-weighted by  $A$*  if there exists a max-weighted path from  $j$  to  $i$  which contains a node of  $A$ . We denote by  $\text{de}_{\text{mw}}^A(j)$  the set of the descendants of  $j$  which are max-weighted by  $A$ . By  $\text{de}_{\text{nmw}}^A(j)$  we further denote the set of the descendants of  $j$  which are *not max-weighted* by  $A$ ; i.e.,  $\text{de}_{\text{nmw}}^A(j) = \text{de}(j) \setminus \text{de}_{\text{mw}}^A(j)$ .  $\square$

Obviously, if  $i \in A$ , then all ancestors and descendants of  $i$  are max-weighted by  $A$ ; i.e., we have  $\text{an}_{\text{mw}}^A(i) = \text{an}(i)$  and  $\text{de}_{\text{mw}}^A(i) = \text{de}(i)$ .

The following auxiliary result presents properties of the coefficients  $b_{ji}$  corresponding to the max-linear representation (2.3) with regard to Definition 2.9(a). We will need them throughout the paper.

**Lemma 2.10.** Let  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  be a ML graphical model with ML coefficient matrix  $B = (b_{ij})_{d \times d}$ . Let  $A \subseteq V$ , and let  $i \in V$  and  $j \in \text{an}(i)$ . Then  $j \in \text{an}_{\text{mw}}^A(i)$  if and only if

$$b_{ji} = \bigvee_{k \in \text{De}(j) \cap A \cap \text{An}(i)} \frac{b_{jk} b_{ki}}{b_{kk}}, \quad (2.6)$$

and  $j \in \text{an}_{\text{nmw}}^A(i)$  if and only if

$$b_{ji} > \bigvee_{k \in \text{De}(j) \cap A \cap \text{An}(i)} \frac{b_{jk} b_{ki}}{b_{kk}}. \quad (2.7)$$

*Proof.* Denote by  $P_{ji,A}$  the set of all paths from  $j$  to  $i$  which pass through at least one node of  $A$ . Observe from (2.2), Definition 2.6, and Definition 2.9(a) that  $j \in \text{an}_{\text{mw}}^A(i)$  if and only if  $b_{ji} = \bigvee_{p \in P_{ji,A}} d_{ji}^p$  and  $j \in \text{an}_{\text{nmw}}^A(i)$  if and only if  $b_{ji} > \bigvee_{p \in P_{ji,A}} d_{ji}^p$ . In analogy to  $P_{ji,A}$  we define  $P_{ji,\{k\}}$  as the set of paths from  $j$  to  $i$  which pass through  $k \in V$ . Since every path from  $j$  to  $i$  in  $P_{ji,A}$  contains some  $k \in A$ , we obtain by (2.2),

$$\bigvee_{p \in P_{ji,A}} d_{ji}^p = \bigvee_{k \in \text{De}(j) \cap A \cap \text{An}(i)} \bigvee_{p \in P_{ji,\{k\}}} d_{ji}^p = \bigvee_{k \in \{j\} \cap A} b_{ki} \vee \bigvee_{k \in \text{de}(j) \cap A \cap \text{An}(i)} \bigvee_{p \in P_{ji,\{k\}}} d_{ji}^p \vee \bigvee_{k \in A \cap \{i\}} b_{jk}.$$

For  $k \in \text{de}(j) \cap A \cap \text{An}(i)$  we can decompose every path from  $j$  to  $i$  containing  $k$  into two successive parts, the first one from  $j$  to  $k$ , and the second from  $k$  to  $i$ . By (2.1) and (2.2) this implies that

$$\bigvee_{p \in P_{ji,\{k\}}} d_{ji}^p = \frac{1}{c_k^k} \left( \bigvee_{p \in P_{jk}} d_{jk}^p \right) \left( \bigvee_{p \in P_{ki}} d_{ki}^p \right).$$



Thus we obtain by (2.2),

$$\bigvee_{k \in \text{de}(j) \cap A \cap \text{an}(i)} \bigvee_{p \in P_{ji, \{k\}}} d_{ji}^p = \bigvee_{k \in \text{de}(j) \cap A \cap \text{an}(i)} \frac{1}{c_k^k} \left( \bigvee_{p \in P_{jk}} d_{jk}^p \right) \left( \bigvee_{p \in P_{ki}} d_{ki}^p \right) = \bigvee_{k \in \text{de}(j) \cap A \cap \text{an}(i)} \frac{b_{jk} b_{ki}}{b_{kk}}.$$

Then altogether we have

$$\bigvee_{p \in P_{ji, A}} d_{ji}^p = \bigvee_{k \in \{j\} \cap A} b_{ki} \vee \bigvee_{k \in \text{de}(j) \cap A \cap \text{an}(i)} \frac{b_{jk} b_{ki}}{b_{kk}} \vee \bigvee_{k \in A \cap \{i\}} b_{jk} = \bigvee_{k \in \text{De}(j) \cap A \cap \text{an}(i)} \frac{b_{jk} b_{ki}}{b_{kk}},$$

and (2.6) as well as (2.7) for  $j \in \text{an}_{\text{nmw}}^A(i)$  follow immediately.  $\square$

For some ML graphical model  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  let  $\mathbf{X}$  be defined on the probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ . For  $\omega \in \Omega$  and  $i \in V$  we denote by  $j(i)$  the node of  $\text{An}(i)$  such that

$$X_i(\omega) = b_{j(i)i} Z_{j(i)}(\omega). \quad (2.8)$$

Note that we have already used this notation in the proof of Proposition 2.3. Obviously,  $j(i)$  depends on  $\omega$  and is therefore random. In the following we present some properties of the node  $j(i)$  and their consequences for the representation of  $\mathbf{X}$ . We show that equality between appropriately scaled components may occur with positive probability, although the corresponding noise variables are continuous. In the context of max-stable processes Dombry et al. [3] call these events *extremal concurrent*. In their Example 2 the extremal concurrence probabilities for max-linear models with unit Fréchet noise variables are calculated.

**Theorem 2.11.** *Let  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  be a ML graphical model with ML coefficient matrix  $B = (b_{ij})_{d \times d}$ . Denote by  $(\Omega, \mathcal{A}, \mathbb{P})$  the probability space on which  $\mathbf{X}$  is defined.*

- (a) *For  $i \in V$  the node  $j(i)$  is  $\mathbb{P}$ -a.s. unique.*
- (b) *For  $i_1, i_2 \in V$  there exists some  $a \in \mathbb{R}_+$  such that the equality  $X_{i_1} = aX_{i_2}$  holds with positive probability if and only if  $\text{An}(i_1) \cap \text{An}(i_2) \neq \emptyset$ . In this case,  $a = \frac{b_{ji_1}}{b_{ji_2}}$  for some  $j \in \text{An}(i_1) \cap \text{An}(i_2)$  and  $\mathbb{P}$ -a.s. on  $\{\omega \in \Omega : X_{i_1}(\omega) = aX_{i_2}(\omega)\}$ ,*

$$j(i_1) = j(i_2) \in M := \left\{ l \in \text{An}(i_1) \cap \text{An}(i_2) : a = \frac{b_{li_1}}{b_{li_2}} \right\};$$

*i.e.,  $X_{i_1} = \bigvee_{l \in M} b_{li_1} Z_l$  and  $X_{i_2} = \bigvee_{l \in M} b_{li_2} Z_l$ .*

*Proof.* (a) Since  $X_i = \bigvee_{j \in \text{An}(i)} b_{ji} Z_j$  and the noise variables are independent and continuous, no two of the scaled noise variables can be equal with positive probability such that the maximum has to be taken  $\mathbb{P}$ -a.s. for a unique node of  $\text{An}(i)$ .

(b) For  $a \in \mathbb{R}_+$  assume that  $\Omega_1 = \{\omega \in \Omega : X_{i_1}(\omega) = aX_{i_2}(\omega)\}$  is non-empty. Then we have on  $\Omega_1$  that

$$X_{i_1} = \bigvee_{l \in \text{An}(i_1)} b_{li_1} Z_l = \bigvee_{l \in \text{An}(i_2)} ab_{li_2} Z_l = aX_{i_2}.$$

Observe from this that as in (a),  $\mathbb{P}(\Omega_1) > 0$  may only hold if  $\text{An}(i_1) \cap \text{An}(i_2) \neq \emptyset$  and  $b_{ji_1} = ab_{ji_2}$  for some  $j \in \text{An}(i_1) \cap \text{An}(i_2)$ .

To prove the converse assume that  $\text{An}(i_1) \cap \text{An}(i_2) \neq \emptyset$  and set  $a = \frac{b_{ji_1}}{b_{ji_2}}$  for some  $j \in \text{An}(i_1) \cap \text{An}(i_2)$ . Note that  $j \in M$  and, hence,  $M \neq \emptyset$ . We obtain on  $\Omega_1$  that

$$X_{i_1} = \bigvee_{l \in M} b_{li_1} Z_l \vee \bigvee_{l \in \text{An}(i_1) \setminus M} b_{li_1} Z_l = \bigvee_{l \in M} b_{li_1} Z_l \vee \frac{b_{ji_1}}{b_{ji_2}} \bigvee_{l \in \text{An}(i_2) \setminus M} b_{li_2} Z_l = aX_{i_2}. \quad (2.9)$$

This holds for instance on

$$\Omega_2 = \left\{ \omega \in \Omega : \bigvee_{l \in M} b_{li_1} Z_l(\omega) > \bigvee_{l \in \text{An}(i_1) \setminus M} b_{li_1} Z_l(\omega) \vee \bigvee_{l \in \text{An}(i_2) \setminus M} ab_{li_2} Z_l(\omega) \right\}.$$

Since the noise variables  $Z_i$  are independent continuous random variables with  $\text{supp}(Z_i) = \mathbb{R}_+$  for  $i = 1, \dots, d$ , we know that  $P(\Omega_2) > 0$  and, as  $\Omega_2 \subseteq \Omega_1$ , also  $P(\Omega_1) > 0$ . Furthermore, we see from (2.9) that, again as in (a),  $X_{i_1}$  and  $X_{i_2}$  are realized  $\mathbb{P}$ -a.s. on  $\Omega_1$  for the same (unique) node in  $M$ ; i.e.,  $j(i_1) = j(i_2) \in M$  and  $X_{i_1} = \bigvee_{l \in M} b_{li_1} Z_l$  as well as  $X_{i_2} = \bigvee_{l \in M} b_{li_2} Z_l$ .  $\square$



**Corollary 2.12.** For  $i \in V$  and  $k \in \text{pa}(i)$  the equality  $X_i = c_k^i X_k$  holds with positive probability if and only if the edge  $[k \rightarrow i]$  is a max-weighted path from  $k$  to  $i$ , which is equivalent to  $b_{ki} = c_k^i c_k^i$ .

*Proof.* By Theorem 2.11(b) the equality  $X_i = c_k^i X_k$  holds with positive probability if and only if  $c_k^i = \frac{b_{ji}}{b_{jk}}$  for some  $j \in \text{An}(k) \cap \text{An}(i) = \text{An}(k)$ . We show that this is equivalent to  $b_{ki} = c_k^i c_k^i$ . If  $b_{ki} = c_k^i c_k^i$ , then  $j = k$ , as  $k \in \text{An}(k)$  and  $b_{kk} = c_k^k$ . To prove the converse assume that  $b_{ji} = b_{jk} c_k^i$  for some  $j \in \text{An}(k) \cap \text{An}(i) = \text{An}(k)$ . We find from Definition 2.6 that in this case there exists a max-weighted path from  $j$  to  $i$  consisting of a max-weighted path from  $j$  to  $k$  and the edge  $[k \rightarrow i]$  and, hence,  $j \in \text{an}_{\text{mw}}^{\{k\}}(i)$ . Then by (2.6) we have  $b_{ji} = \frac{b_{jk} b_{ki}}{b_{kk}}$  and, thus, as  $b_{ji} = b_{jk} c_k^i$ ,  $b_{ki} = b_{kk} c_k^i = c_k^k c_k^i$ .  $\square$

### 3. Characterization of a max-linear graphical model

In Theorem 2.2 we have started with a ML graphical model  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  and have derived the max-linear representation of  $\mathbf{X}$ . In this section, for a given max-linear model (2.4), we investigate under which conditions this model has a representation as in (1.3) or, equivalently, when gives this model rise to a ML graphical model.

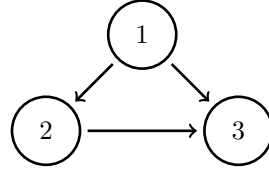
First, recall from Remark 2.4(i) that a DAG corresponding to a ML graphical model with ML coefficient matrix  $B = (b_{ij})_{d \times d}$  has reachability matrix  $R = \text{sgn}(B)$ . Thus this condition has to be satisfied in any case. Before we discuss further conditions we give an example.

**Example 3.1.** [Not every ML model gives rise to a ML graphical model]  
Consider for  $d = 3$  the max-linear model

$$\mathbf{X} = \mathbf{Z} \odot B \quad \text{with} \quad B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ 0 & b_{22} & b_{23} \\ 0 & 0 & b_{33} \end{bmatrix}$$

such that  $b_{ij} > 0$  for  $i \leq j$ . Thus  $B$  is a weighted reachability matrix corresponding to the DAG given by

$$\mathcal{D} = (\{1, 2, 3\}, \{(1, 2), (1, 3), (2, 3)\}).$$



Assume that  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  is a ML graphical model. Thus we have

$$\begin{aligned} X_1 &= b_{11} Z_1 = c_1^1 Z_1 \\ X_2 &= b_{12} Z_1 \vee b_{22} Z_2 = c_1^2 X_1 \vee c_2^2 Z_2 = c_1^2 b_{11} Z_1 \vee c_2^2 Z_2 \\ X_3 &= b_{13} Z_1 \vee b_{23} Z_2 \vee b_{33} Z_3 = c_1^3 X_1 \vee c_2^3 X_2 \vee c_3^3 Z_3 \\ &= c_1^3 b_{11} Z_1 \vee c_2^3 (b_{12} Z_1 \vee b_{22} Z_2) \vee b_{33} Z_3 = (c_1^3 b_{11} \vee c_2^3 b_{12}) Z_1 \vee c_2^3 b_{22} Z_2 \vee c_3^3 Z_3. \end{aligned}$$

Hence,

$$c_1^1 = b_{11}, \quad c_1^2 = \frac{b_{12}}{b_{11}}, \quad c_2^2 = b_{22}, \quad c_2^3 = \frac{b_{23}}{b_{22}}, \quad c_3^3 = b_{33}, \quad \text{and} \quad b_{13} = c_1^3 b_{11} \vee c_2^3 b_{12} = c_1^3 b_{11} \vee \frac{b_{12} b_{23}}{b_{22}}.$$

In order to find  $c_1^3$  we consider the three possible cases:

- (1) If  $b_{13} < \frac{b_{12} b_{23}}{b_{22}}$ , then there is no DAG consistent with  $B$ ; i.e.,  $B$  is not a ML coefficient matrix corresponding to a ML graphical model.
- (2) If  $b_{13} > \frac{b_{12} b_{23}}{b_{22}}$ , then  $c_1^3 = \frac{b_{13}}{b_{11}}$  is consistent with  $\mathcal{D}$ , and  $B$  is a ML coefficient matrix corresponding to a ML graphical model.
- (3) If  $b_{13} = \frac{b_{12} b_{23}}{b_{22}}$ , then every  $c_1^3 < \frac{b_{13}}{b_{11}}$  is consistent with  $\mathcal{D}$ . This indicates that the edge  $[1 \rightarrow 3]$  is irrelevant and can be dropped. Thus  $(\mathcal{D}_1, \mathcal{L}(\mathbf{X}))$  with  $\mathcal{D}_1 = (\{1, 2, 3\}, \{(1, 2), (2, 3)\})$  is also a ML graphical model, and the matrix  $B$  is also a ML coefficient matrix corresponding to a ML graphical model.

□

Observe that in Example 3.1 the DAG  $(\{1, 2, 3\}, \{(1, 2), (2, 3)\})$  is the smallest DAG (with respect to the number of edges) such that  $\text{sgn}(B)$  is its reachability matrix. We give the general definition of this specific DAG, see for example Aho et al. [1].

**Definition 3.2.** Let  $\mathcal{D} = (V, E)$  be a DAG. A graph  $\mathcal{D}^{\text{tr}} = (V, E^{\text{tr}})$  is a *transitive reduction* of  $\mathcal{D}$  if

- (a) for all  $i, j \in V$ ,  $\mathcal{D}^{\text{tr}}$  has a path from  $j$  to  $i$  if and only if  $\mathcal{D}$  has a path from  $j$  to  $i$ , and
- (b) there is no graph with fewer edges than  $\mathcal{D}^{\text{tr}}$  satisfying condition (a).

□

Since we work with finite DAGs throughout, the transitive reduction is unique and is also a subgraph of the original DAG. For a finite DAG it is therefore the same as the so-called *minimum equivalent graph* (see for example Moyles and Thompson [7]).

The transitive reduction  $\mathcal{D}^{\text{tr}}$  of a DAG  $\mathcal{D}$  can be computed from the original graph by successively examining the edges of  $\mathcal{D}$ , in any order, and deleting those edges which are redundant, where an edge  $[j \rightarrow i]$  is redundant if the graph contains a path from  $j$  to  $i$  which does not include this edge. Obviously,  $\mathcal{D}$  and  $\mathcal{D}^{\text{tr}}$  have the same reachability matrix. We denote by  $\text{pa}^{\text{tr}}(i)$  the set of parents of  $i$  with respect to the transitive reduction of  $\mathcal{D}$ .

The following theorem characterizes all max-linear random vectors which give rise to a ML graphical model.

**Theorem 3.3.** Let  $\mathcal{D}$  be a DAG with reachability matrix  $R$ . Assume that  $\mathbf{X}$  is a ML model as in (2.4) with ML coefficient matrix  $B = (b_{ij})_{d \times d}$  such that  $R = \text{sgn}(B)$ . Define

$$D = (d_{ij})_{d \times d} := \text{diag}(b_{ii}, i = 1, \dots, d) \quad \text{and} \quad D_0 = (d_{0,ji})_{d \times d} := \left( \frac{b_{ji}}{b_{jj}} \mathbf{1}([j \rightarrow i]) \right)_{d \times d}.$$

Then  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  is a ML graphical model if and only if the following fixed point equation holds:

$$B = D \vee B \odot D_0. \quad (3.1)$$

In this case,

$$X_i = \bigvee_{k \in \text{pa}(i)} \frac{b_{ki}}{b_{kk}} X_k \vee b_{ii} Z_i;$$

i.e., the weights in (1.3) are given by  $c_k^i = \frac{b_{ki}}{b_{kk}}$  and  $c_i^i = b_{ii}$ .

*Proof.* Let  $\mathcal{D}^{\text{tr}}$  be the transitive reduction of  $\mathcal{D}$ . First, we have a closer look at the fixed point equation (3.1). We start computing the right-hand side of (3.1) for the  $ji$ -th component and obtain

$$d_{ji} \vee \bigvee_{k=1}^d \frac{b_{jk} b_{ki}}{b_{kk}} \mathbf{1}([k \rightarrow i]) = d_{ji} \vee \bigvee_{k \in \text{De}(j) \cap \text{pa}(i)} \frac{b_{jk} b_{ki}}{b_{kk}}, \quad (3.2)$$

since  $R = \text{sgn}(B)$ . For  $j \notin \text{an}(i)$  we have  $\text{De}(j) \cap \text{pa}(i) = \emptyset$  such that (3.2) reduces to the coefficient in  $D$ . For  $j \in \text{an}(i)$  we have  $d_{ji} = 0$  such that (3.2) reduces to the coefficient in  $B \odot D_0$ . Furthermore, if  $j \in \text{an}(i)$ , then either  $j \in \text{an}(i) \setminus \text{pa}(i)$ ,  $j \in \text{pa}(i) \setminus \text{pa}^{\text{tr}}(i)$ , or  $j \in \text{pa}^{\text{tr}}(i)$ .

For  $j \in \text{an}(i) \setminus \text{pa}(i)$  (3.2) reduces to

$$\bigvee_{k \in \text{de}(j) \cap \text{pa}(i)} \frac{b_{jk} b_{ki}}{b_{kk}}.$$

Note that  $\text{de}(j) \cap \text{pa}(i) \neq \emptyset$  for  $j \in \text{pa}(i) \setminus \text{pa}^{\text{tr}}(i)$  and  $\text{de}(j) \cap \text{pa}(i) = \emptyset$  for  $j \in \text{pa}^{\text{tr}}(i)$ . Thus, for  $j \in \text{pa}(i) \setminus \text{pa}^{\text{tr}}(i)$  we obtain for (3.2),

$$\bigvee_{k \in \text{De}(j) \cap \text{pa}(i)} \frac{b_{jk} b_{ki}}{b_{kk}} = b_{ji} \vee \bigvee_{k \in \text{de}(j) \cap \text{pa}(i)} \frac{b_{jk} b_{ki}}{b_{kk}},$$

and for  $j \in \text{pa}^{\text{tr}}(i)$  (3.2) reduces to

$$\bigvee_{k \in \text{de}(j) \cap \text{pa}(i)} \frac{b_{jk}b_{ki}}{b_{kk}} = b_{ji}.$$

Altogether we obtain that (3.1) holds if and only if for  $i, j = 1, \dots, d$

$$b_{ji} = \begin{cases} 0, & \text{if } j \notin \text{An}(i), \\ b_{ii}, & \text{if } j = i, \\ \bigvee_{k \in \text{de}(j) \cap \text{pa}(i)} \frac{b_{jk}b_{ki}}{b_{kk}}, & \text{if } j \in \text{an}(i) \setminus \text{pa}(i), \\ b_{ji} \vee \bigvee_{k \in \text{de}(j) \cap \text{pa}(i)} \frac{b_{jk}b_{ki}}{b_{kk}}, & \text{if } j \in \text{pa}(i) \setminus \text{pa}^{\text{tr}}(i), \\ b_{ji}, & \text{if } j \in \text{pa}^{\text{tr}}(i). \end{cases}$$

Since for  $j \notin \text{An}(i)$ ,  $R = \text{sgn}(B)$  implies  $b_{ji} = 0$ , the fixed point equation (3.1) is satisfied if and only if

$$b_{ji} = \bigvee_{k \in \text{de}(j) \cap \text{pa}(i)} \frac{b_{jk}b_{ki}}{b_{kk}} \quad \text{for all } j \in \text{an}(i) \setminus \text{pa}(i) \quad (3.3)$$

and

$$b_{ji} = b_{ji} \vee \bigvee_{k \in \text{de}(j) \cap \text{pa}(i)} \frac{b_{jk}b_{ki}}{b_{kk}} \quad \text{for all } j \in \text{pa}(i) \setminus \text{pa}^{\text{tr}}(i). \quad (3.4)$$

Now we prove that, under the conditions above,  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  is a ML graphical model if and only if (3.1) or, equivalently, (3.3) and (3.4) hold.

Assume that  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  is a ML graphical model. For  $j \in \text{an}(i)$  observe that  $j \in \text{an}_{\text{mw}}^{\text{pa}(i)}(i)$ , since every path from  $j$  to  $i$  contains a node of  $\text{pa}(i)$ . Thus by (2.6) we have for  $j \in \text{an}(i) \setminus \text{pa}(i)$ ,

$$b_{ji} = \bigvee_{k \in \text{de}(j) \cap \text{pa}(i) \cap \text{an}(i)} \frac{b_{jk}b_{ki}}{b_{kk}} = \bigvee_{k \in \text{de}(j) \cap \text{pa}(i)} \frac{b_{jk}b_{ki}}{b_{kk}}.$$

and for  $j \in \text{pa}(i) \setminus \text{pa}^{\text{tr}}(i)$ ,

$$b_{ji} = \bigvee_{k \in \text{de}(j) \cap \text{pa}(i) \cap \text{An}(i)} \frac{b_{jk}b_{ki}}{b_{kk}} = b_{ji} \vee \bigvee_{k \in \text{de}(j) \cap \text{pa}(i)} \frac{b_{jk}b_{ki}}{b_{kk}}.$$

Assume conversely that (3.3) and (3.4) hold. We split up the index set and use (3.3) in the first place, and (3.4) in the second place to obtain

$$\begin{aligned} X_i &= \bigvee_{j \in \text{An}(i)} b_{ji} Z_j \\ &= \bigvee_{j \in \text{an}(i) \setminus \text{pa}(i)} b_{ji} Z_j \vee \bigvee_{j \in \text{pa}(i) \setminus \text{pa}^{\text{tr}}(i)} b_{ji} Z_j \vee \bigvee_{j \in \text{pa}^{\text{tr}}(i)} b_{ji} Z_j \vee b_{ii} Z_i \\ &= \bigvee_{j \in \text{an}(i) \setminus \text{pa}(i)} \bigvee_{k \in \text{de}(j) \cap \text{pa}(i)} \frac{b_{jk}b_{ki}}{b_{kk}} Z_j \vee \bigvee_{j \in \text{pa}(i) \setminus \text{pa}^{\text{tr}}(i)} (b_{ji} \vee \bigvee_{k \in \text{de}(j) \cap \text{pa}(i)} \frac{b_{jk}b_{ki}}{b_{kk}}) Z_j \vee \bigvee_{j \in \text{pa}^{\text{tr}}(i)} b_{ji} Z_j \vee b_{ii} Z_i. \end{aligned}$$

Now we use (A.2) and (A.3) to interchange the first two as well as the fourth and the fifth maxima to obtain

$$X_i = \bigvee_{k \in \text{pa}(i)} \bigvee_{j \in \text{an}(k) \setminus \text{pa}(i)} \frac{b_{jk}b_{ki}}{b_{kk}} Z_j \vee \bigvee_{j \in \text{pa}(i) \setminus \text{pa}^{\text{tr}}(i)} b_{ji} Z_j \vee \bigvee_{k \in \text{pa}(i)} \bigvee_{j \in \text{an}(k) \cap \text{pa}(i) \setminus \text{pa}^{\text{tr}}(i)} \frac{b_{jk}b_{ki}}{b_{kk}} Z_j \vee \bigvee_{j \in \text{pa}^{\text{tr}}(i)} b_{ji} Z_j \vee b_{ii} Z_i.$$

Next, observe for  $k \in \text{pa}(i)$  that  $\text{an}(k) \cap \text{pa}(i) \setminus \text{pa}^{\text{tr}}(i) = \text{an}(k) \cap \text{pa}(i)$ . Indeed, assume some  $l \in \text{pa}^{\text{tr}}(i)$  such that  $l \in \text{an}(k) \cap \text{pa}(i)$ . Then,  $k \in \text{de}(l) \cap \text{pa}(i)$  which implies that  $\text{de}(l) \cap \text{pa}(i) \neq \emptyset$ . This is a contradiction

to the fact that  $l \in \text{pa}^{\text{tr}}(i)$  (since  $\text{de}(k) \cap \text{pa}(i) = \emptyset$  for all  $k \in \text{pa}^{\text{tr}}(i)$ ). Moreover, we combine the maxima over  $j \in \text{pa}(i) \setminus \text{pa}^{\text{tr}}(i)$  and  $j \in \text{pa}^{\text{tr}}(i)$ . Thus we obtain

$$\begin{aligned}
X_i &= \bigvee_{k \in \text{pa}(i)} \frac{b_{ki}}{b_{kk}} \bigvee_{j \in \text{an}(k) \setminus \text{pa}(i)} b_{jk} Z_j \vee \bigvee_{j \in \text{pa}(i)} b_{ji} Z_j \vee \bigvee_{k \in \text{pa}(i)} \frac{b_{ki}}{b_{kk}} \bigvee_{j \in \text{an}(k) \cap \text{pa}(i)} b_{jk} Z_j \vee b_{ii} Z_i \\
&= \bigvee_{k \in \text{pa}(i)} \left( \frac{b_{ki}}{b_{kk}} \bigvee_{j \in \text{an}(k) \setminus \text{pa}(i)} b_{jk} Z_j \vee b_{ki} Z_k \vee \frac{b_{ki}}{b_{kk}} \bigvee_{j \in \text{an}(k) \cap \text{pa}(i)} b_{jk} Z_j \right) \vee b_{ii} Z_i \\
&= \bigvee_{k \in \text{pa}(i)} \left( \frac{b_{ki}}{b_{kk}} \bigvee_{j \in \text{an}(k)} b_{jk} Z_j \vee b_{ki} Z_k \right) \vee b_{ii} Z_i \\
&= \bigvee_{k \in \text{pa}(i)} \frac{b_{ki}}{b_{kk}} \left( \bigvee_{j \in \text{an}(k)} b_{jk} Z_j \vee b_{kk} Z_k \right) \vee b_{ii} Z_i \\
&= \bigvee_{k \in \text{pa}(i)} \frac{b_{ki}}{b_{kk}} X_k \vee b_{ii} Z_i.
\end{aligned}$$

Hence,  $\mathbf{X}$  has representation (1.3) with weights  $c_k^i = \frac{b_{ki}}{b_{kk}}$  and  $c_i^i = b_{ii}$ , and  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  is a ML graphical model.  $\square$

In the proof of Theorem 3.3 we have shown that under the given conditions the fixed point equation (3.1) holds if and only if (3.3) and (3.4) hold. This implies the following equivalent version of Theorem 3.3.

**Corollary 3.4.** *Let  $\mathcal{D}^{\text{tr}}$  be the transitive reduction of  $\mathcal{D}$ . Then  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  is a ML graphical model if and only if for every  $i = 1, \dots, d$ ,*

$$b_{ji} = \bigvee_{k \in \text{de}(j) \cap \text{pa}(i)} \frac{b_{jk} b_{ki}}{b_{kk}} \quad \text{for all } j \in \text{an}(i) \setminus \text{pa}(i) \quad (3.5)$$

and

$$b_{ji} \geq \bigvee_{k \in \text{de}(j) \cap \text{pa}(i)} \frac{b_{jk} b_{ki}}{b_{kk}} \quad \text{for all } j \in \text{pa}(i) \setminus \text{pa}^{\text{tr}}(i).$$

Given a DAG  $\mathcal{D}$  Theorem 3.3 or Corollary 3.4 allow us to find all ML models  $\mathbf{X}$  such that  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  is a ML graphical model. On the other hand, we know from Example 3.1 that for a ML graphical model  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  the distribution  $\mathcal{L}(\mathbf{X})$  may also form together with other DAGs a ML graphical model. This raises the question what the smallest DAG (with respect to the number of edges) is, forming together with  $\mathcal{L}(\mathbf{X})$  a ML graphical model. The following provides an answer.

**Theorem 3.5.** *Let  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  be a ML graphical model with ML coefficient matrix  $B = (b_{ij})_{d \times d}$ . Let further  $\mathcal{D}^{\text{tr}} = (V, E^{\text{tr}})$  be the transitive reduction of  $\mathcal{D}$ . Define*

$$\begin{aligned}
B^= &:= \left\{ (k, i) \in V \times V : k \in \text{pa}(i) \setminus \text{pa}^{\text{tr}}(i) \cap \text{an}_{\text{mw}}^{\text{pa}(i) \setminus \{k\}}(i) \right\} \\
&= \left\{ (k, i) \in V \times V : k \in \text{pa}(i) \setminus \text{pa}^{\text{tr}}(i) \text{ and } b_{ki} = \bigvee_{l \in \text{de}(k) \cap \text{pa}(i)} \frac{b_{kl} b_{li}}{b_{ll}} \right\}
\end{aligned}$$

and for  $E^B := E \setminus B^=$  the DAG  $\mathcal{D}^B := (V, E^B)$ . Then  $\mathcal{D}^B$  is the DAG with minimal number of edges such that  $(\mathcal{D}^B, \mathcal{L}(\mathbf{X}))$  is a ML graphical model.

*Proof.* The equality of the two sets in the definition of  $B^=$  follows from (2.6). First, we prove that  $(\mathcal{D}^B, \mathcal{L}(\mathbf{X}))$  is a ML graphical model. By Theorem 2.2 and Definition 2.6 it suffices to show that for all  $i \in V$  and  $j \in \text{an}(i)$  there exists a max-weighted path which only contains edges of  $E^B$ . For some  $i \in V$  and  $j \in \text{an}(i)$  let  $p = [j = k_0 \rightarrow k_1 \rightarrow \dots \rightarrow k_n = i]$  be a max-weighted path in  $\mathcal{D}$  with maximal number of edges of all max-weighted paths from  $j$  to  $i$ . Assume that  $p$  contains an edge  $k_{l-1} \rightarrow k_l$  with  $(k_{l-1}, k_l) \in B^=$  for some  $l \in \{1, \dots, n\}$ . By the definition of  $B^=$  there exists a max-weighted path  $p_1$  from  $k_{l-1}$  to  $k_l$  which does not include the edge  $k_{l-1} \rightarrow k_l$ . Thus by replacing in  $p$  the edge  $k_{l-1} \rightarrow k_l$  by  $p_1$

we obtain by Remark 2.7(iv) a max-weighted path from  $j$  to  $i$  containing more edges than  $p$ . This is, however, a contradiction to the fact that  $p$  is of maximal length of all max-weighted paths from  $j$  to  $i$ .

It remains to show that there exists no DAG which has less edges than  $\mathcal{D}^B$  and forms together with  $\mathcal{L}(\mathbf{X})$  a ML graphical model. Recall from Remark 2.4(i) that a DAG corresponding to a ML graphical model cannot have less edges than the transitive reduction of any DAG with reachability matrix  $R = \text{sgn}(B)$ ; i.e.,  $E^{\text{tr}} \subseteq E^B$ . Let  $i \in V$  and  $k \in \text{pa}(i) \setminus \text{pa}^{\text{tr}}(i)$  such that the only max-weighted path from  $k$  to  $i$  is  $[k \rightarrow i]$ ; i.e.,  $k \in \text{an}_{\text{nmw}}^{\text{pa}(i) \setminus \{k\}}(i)$  and, hence,  $k \in E^B \setminus E^{\text{tr}}$ . Since  $[k \rightarrow i]$  is the only max-weighted path from  $k$  to  $i$  in  $\mathcal{D}$ , every path from  $k$  to  $i$  in a subgraph of  $\mathcal{D}$  which does not contain the edge  $k \rightarrow i$  has a smaller weight. Thus all edges of  $E^B$  have to be in a DAG, which forms together with  $\mathcal{L}(\mathbf{X})$  a ML graphical model.  $\square$

We call  $\mathcal{D}^B$  as in Theorem 3.5 *minimal max-linear DAG*, and for  $i \in V$  we denote by  $\text{pa}^B(i)$  the parents of  $i$  with respect to  $\mathcal{D}^B$ .

For  $i \in V$  and  $k \in \text{pa}^B(i)$  observe from the definition of  $\mathcal{D}^B$  that the only max-weighted path from  $k$  to  $i$  is  $[k \rightarrow i]$ . Thus the following is an immediate consequence of Definition 2.6 and Corollary 2.12.

**Corollary 3.6.** *Let  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  be a ML graphical model with ML coefficient matrix  $B = (b_{ij})_{d \times d}$ . Assume that the DAG  $\mathcal{D}$  is minimal max-linear; i.e.,  $\mathcal{D} = \mathcal{D}^B$ . Then for all  $i \in V$  and  $k \in \text{pa}(i)$ ,*

- (a) *the weights  $c_k^i$  are uniquely given by  $B$ , namely,  $c_k^i = \frac{b_{ki}}{b_{kk}}$ , and*
- (b) *the equality  $X_i = c_k^i X_k$  holds with positive probability.*

**Remark 3.7.** There are two advantages for considering  $(\mathcal{D}^B, \mathcal{L}(\mathbf{X}))$  instead of  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$ :

- (i) The representation (1.3) of  $\mathbf{X}$  is unique.
- (ii) As we know from (1.2), missing edges correspond to conditional independence in the distribution of  $\mathbf{X}$ ; i.e., we can read off from  $\mathcal{D}^B$  more conditional independence properties of  $\mathcal{L}(\mathbf{X})$  than from  $\mathcal{D}$ .  $\square$

#### 4. Structural properties of a max-linear graphical model

Throughout this section we assume that  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  is a ML graphical model with ML coefficient matrix  $B = (b_{ij})_{d \times d}$ . From (1.3), we learn immediately that

$$c_k^i X_k \leq X_i \quad \text{for all } k \in \text{pa}(i) \text{ and } i \in V. \quad (4.1)$$

The following lemma states inequality (4.1) for nodes  $X_j$  and  $X_i$  which are not simply connected by an edge  $[j \rightarrow i]$ , but by a path  $[j \Rightarrow i]$  of arbitrary length.

**Lemma 4.1.** *Let  $U \subseteq V$ . Then for  $i \in V$ ,*

$$\bigvee_{j \in \text{an}(i) \cap U} \frac{b_{ji}}{b_{jj}} X_j \leq X_i \leq \bigwedge_{l \in \text{de}(i) \cap U} \frac{b_{il}}{b_{ll}} X_l. \quad (4.2)$$

*Proof.* Assume nodes  $i, j \in V$  such that  $j \in \text{an}(i)$ , and let  $p = [j = k_0 \rightarrow k_1 \rightarrow \dots \rightarrow k_n = i]$  be an arbitrary path in  $\mathcal{D}$ . By applying (4.1) iteratively we obtain

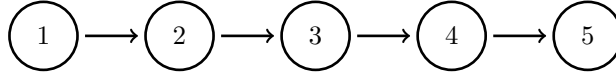
$$c_j^i X_i = c_{k_0}^{k_n} X_{k_n} \geq c_{k_0}^{k_0} c_{k_0}^{k_{n-1}} X_{k_{n-1}} \geq \dots \geq c_{k_0}^{k_0} \prod_{l=0}^{n-1} c_{k_l}^{k_{l+1}} X_{k_0} = d_{ji}^p X_j$$

and, thus, as this holds for all paths,  $c_j^i X_i \geq \bigvee_{p \in P_{ji}} d_{ji}^p X_j$ , which is by (2.2) equivalent to  $\frac{b_{ji}}{b_{jj}} X_j \leq X_i$ . Thus we have  $\frac{b_{ji}}{b_{jj}} X_j \leq X_i$  for all  $j \in \text{an}(i) \cap U$  and  $X_i \leq \frac{b_{il}}{b_{ll}} X_l$  for all  $l \in \text{de}(i) \cap U$ , which, finally, implies (4.2).  $\square$

The bounds given in (4.2) can, under certain conditions, be modified such that they are based on a smaller number of nodes. In certain cases, they reduce to a single properly scaled random variable  $X_j$  and  $X_l$ , respectively. The following example illustrates this.

**Example 4.2.** Consider the ML graphical model  $(\mathcal{D}, \mathcal{L}(X_1, \dots, X_5))$  with DAG

$$\mathcal{D} = (\{1, 2, 3, 4, 5\}, \{(1, 2), (2, 3), (3, 4), (4, 5)\})$$



and ML coefficient matrix  $B = (b_{ij})_{5 \times 5}$  in the context of Lemma 4.1. For  $U = \{1, 2, 4, 5\}$  and  $i = 3$  we obtain

$$\frac{b_{13}}{b_{11}}X_1 \wedge \frac{b_{23}}{b_{22}}X_2 \leq X_3 \leq \frac{b_{33}}{b_{34}}X_4 \wedge \frac{b_{33}}{b_{35}}X_5. \quad (4.3)$$

Furthermore, using Lemma 4.1 with  $U = \{4\}$  and  $i = 5$  as well as (3.5), we obtain

$$\frac{b_{45}}{b_{44}}X_4 \leq X_5 \iff \frac{b_{34}b_{45}}{b_{44}}X_4 \leq b_{34}X_5 \iff b_{35}X_4 \leq b_{34}X_5 \iff \frac{b_{33}}{b_{34}}X_4 \leq \frac{b_{33}}{b_{35}}X_5.$$

Similarly, we derive  $\frac{b_{13}}{b_{11}}X_1 \leq \frac{b_{23}}{b_{22}}X_2$ . Thus we have identified the bounds in (4.3) as

$$\frac{b_{13}}{b_{11}}X_1 \wedge \frac{b_{23}}{b_{22}}X_2 = \frac{b_{23}}{b_{22}}X_2 \quad \text{and} \quad \frac{b_{33}}{b_{34}}X_4 \wedge \frac{b_{33}}{b_{35}}X_5 = \frac{b_{33}}{b_{34}}X_4.$$

□

In order to investigate this effect in general, we need the following definition. Since it is not completely obvious how to obtain these particular ancestors and descendants, we present an algorithm in Remark 4.4 and illustrate the concept in Example 4.5 below.

**Definition 4.3.** Let  $U \subseteq V$ .

- (a) For nodes  $i, j \in V$  such that  $j \in \text{an}(i) \cap U$ , we call  $j$  *lowest max-weighted ancestor of  $i$  in  $U$*  if  $U \setminus \{i, j\} = \emptyset$  or if  $U \setminus \{i, j\} \neq \emptyset$  and there exists no max-weighted path from  $j$  to  $i$  containing nodes of  $U \setminus \{i, j\}$ . We denote the set of the lowest max-weighted ancestors of  $i$  in  $U$  by  $\text{an}_{\text{low}}^U(i)$ .
- (b) For nodes  $i, j \in V$  such that  $i \in \text{de}(j) \cap U$ , we call  $i$  *highest max-weighted descendant of  $j$  in  $U$*  if  $U \setminus \{i, j\} = \emptyset$  or if  $U \setminus \{i, j\} \neq \emptyset$  and there exists no max-weighted path from  $j$  to  $i$  containing nodes of  $U \setminus \{i, j\}$ . We denote the set of the highest max-weighted descendants of  $j$  in  $U$  by  $\text{de}_{\text{high}}^U(j)$ . □

**Remark 4.4.** Assume the situation of Definition 4.3(a). The set  $\text{an}_{\text{low}}^U(i)$  may be determined as follows:

- (i) Identify every  $j \in \text{an}(i) \cap U$  such that  $U \setminus \{i, j\} = \emptyset$  or that  $U \setminus \{i, j\} \neq \emptyset$  and there exists a path  $[j \Rightarrow i]$  which does not contain any node of  $U \setminus \{i, j\}$ . We summarize these nodes in the set  $A$ .
- (ii) For all  $j \in A$  such that  $U \setminus \{i, j\} \neq \emptyset$ , remove  $j$  from  $A$  if there exists a max-weighted path  $[j \Rightarrow i]$  containing some node of  $U \setminus \{i, j\}$ .
- (iii) The remaining nodes of  $A$  are the lowest max-weighted ancestors of  $i$  in  $U$ ; i.e.,  $A = \text{an}_{\text{low}}^U(i)$ . □

**Example 4.5.** [Continuation of Examples 2.1 and 2.8:  $\text{an}_{\text{low}}^U(i)$  and  $\text{de}_{\text{high}}^U(j)$ ]

Consider  $U = \{1, 2\}$  and  $i = 4$  in the context of Definition 4.3(a). The only and thus max-weighted path  $[2 \rightarrow 4]$  from 2 to 4 does not contain any node of  $U \setminus \{2, 4\} = \{1\}$ ; i.e.,  $2 \in \text{an}_{\text{low}}^U(4)$ . It remains to discuss node 1, which gives rise to two cases:

- (1) If  $b_{14} = \frac{b_{12}b_{24}}{b_{22}} \geq \frac{b_{13}b_{34}}{b_{33}}$ , then the max-weighted path  $[1 \rightarrow 2 \rightarrow 4]$  does also contain node 2 of  $U \setminus \{1, 4\} = \{2\}$ . Thus we have that  $1 \notin \text{an}_{\text{low}}^U(4)$  and, consequently,  $\text{an}_{\text{low}}^U(4) = \{2\}$ .
- (2) If  $b_{14} = \frac{b_{13}b_{34}}{b_{33}} > \frac{b_{12}b_{24}}{b_{22}}$ , then the only max-weighted path from 1 to 4 is  $[1 \rightarrow 3 \rightarrow 4]$ . Since this path does not contain node 2, we have  $\text{an}_{\text{low}}^U(4) = \{1, 2\}$ .

Now consider the case  $U = \{3, 4\}$  and  $j = 1$  in the context of Definition 4.3(b). Similarly as above, we obtain the two cases:

- (1) If  $b_{14} = \frac{b_{13}b_{34}}{b_{33}} \geq \frac{b_{12}b_{24}}{b_{22}}$ , then  $\text{de}_{\text{high}}^U(1) = \{3\}$ .  
(2) If  $b_{14} = \frac{b_{12}b_{24}}{b_{22}} > \frac{b_{13}b_{34}}{b_{33}}$ , then  $\text{de}_{\text{high}}^U(1) = \{3, 4\}$ .  $\square$

The following shows now that, in order to determine the lower and upper bounds in (4.2), it suffices to consider the lowest max-weighted ancestors of  $i$  in  $U$  and the highest max-weighted descendants of  $i$  in  $U$ .

**Lemma 4.6.** *Let  $U \subseteq V$ . Then for  $i \in V$ ,*

$$\bigvee_{j \in \text{an}_{\text{low}}^U(i)} \frac{b_{ji}}{b_{jj}} X_j = \bigvee_{j \in \text{an}(i) \cap U} \frac{b_{ji}}{b_{jj}} X_j \leq X_i \leq \bigwedge_{l \in \text{de}_{\text{high}}^U(i)} \frac{b_{ii}}{b_{il}} X_l = \bigwedge_{l \in \text{de}(i) \cap U} \frac{b_{ii}}{b_{il}} X_l. \quad (4.4)$$

*Proof.* Since (4.2) holds, we only have to show the equalities for the lower and upper bounds. Let  $k \in (\text{an}(i) \cap U) \setminus \text{an}_{\text{low}}^U(i)$ . By the definition of the set  $\text{an}_{\text{low}}^U(i)$  we know that  $U \setminus \{k, i\} \neq \emptyset$  and that there exists a max-weighted path from  $k$  to  $i$  containing some node of  $U \setminus \{k, i\}$ ; i.e., there exists a max-weighted path containing some node of  $U \setminus \{i\}$ . Then by Lemma A.2(a) there exists a max-weighted path from  $k$  to  $i$  containing some  $j \in \text{an}_{\text{low}}^U(i)$ ; i.e.,  $k \in \text{an}_{\text{mw}}^{\{j\}}(i)$ . Thus, using (2.6) and Lemma 4.1, we obtain

$$\frac{b_{ki}}{b_{kk}} X_k = \frac{b_{kj}b_{ji}}{b_{kk}b_{jj}} X_k \leq \frac{b_{ji}}{b_{jj}} X_j. \quad (4.5)$$

Since for all  $k \in (\text{an}(i) \cap U) \setminus \text{an}_{\text{low}}^U(i)$ , there exists some  $j \in \text{an}_{\text{low}}^U(i)$  such that (4.5) holds, we have

$$\bigvee_{j \in \text{an}(i) \cap U} \frac{b_{ji}}{b_{jj}} X_j = \bigvee_{j \in \text{an}_{\text{low}}^U(i)} \frac{b_{ji}}{b_{jj}} X_j.$$

Similarly, we obtain the equality for the upper bound of  $X_i$  in (4.4).  $\square$

Next, for  $U \subsetneq V$ , we write the random variables  $X_i$  for  $i \in U^c := V \setminus U$  as functions of the random vector  $\mathbf{X}_U$  and noise variables. Of course, there are many such representations. We focus on that with minimal number of nodes in  $U$  and noise variables.

**Theorem 4.7.** *Let  $U \subsetneq V$  and denote  $\text{An}_{\text{nmw}}^U(i) = \text{an}_{\text{nmw}}^U(i) \cup \{i\}$  (cf. Definition 2.9(a)). Then for  $i \in U^c$ ,*

$$X_i = \bigvee_{k \in \text{an}_{\text{low}}^U(i)} \frac{b_{ki}}{b_{kk}} X_k \vee \bigvee_{j \in \text{An}_{\text{nmw}}^U(i)} b_{ji} Z_j. \quad (4.6)$$

*This representation of  $X_i$  as a function of  $\mathbf{X}_U$  and noise variables involves the minimal number of nodes in  $U$  and noise variables.*

*Proof.* Using the first equality in (4.4) and (2.3) as well as in a second step (A.4) to interchange the first two maxima, we obtain for the right-hand side of (4.6)

$$\begin{aligned} \bigvee_{k \in \text{an}_{\text{low}}^U(i)} \frac{b_{ki}}{b_{kk}} X_k \vee \bigvee_{j \in \text{An}_{\text{nmw}}^U(i)} b_{ji} Z_j &= \bigvee_{k \in \text{an}(i) \cap U} \frac{b_{ki}}{b_{kk}} \left( \bigvee_{j \in \text{An}(k)} b_{jk} Z_j \right) \vee \bigvee_{j \in \text{An}_{\text{nmw}}^U(i)} b_{ji} Z_j \\ &= \bigvee_{j \in \text{an}(i)} \bigvee_{k \in \text{De}(j) \cap \text{an}(i) \cap U} \frac{b_{jk}b_{ki}}{b_{kk}} Z_j \vee \bigvee_{j \in \text{An}_{\text{nmw}}^U(i)} b_{ji} Z_j. \end{aligned}$$

Splitting the set  $\text{an}(i)$  into  $\text{an}_{\text{mw}}^U(i)$  and  $\text{an}(i) \setminus \text{an}_{\text{mw}}^U(i) = \text{an}_{\text{nmw}}^U(i)$  and, as  $i \in U^c$ , using (2.6) yields

$$\begin{aligned} \bigvee_{j \in \text{an}_{\text{mw}}^U(i)} \bigvee_{k \in \text{De}(j) \cap \text{an}(i) \cap U} \frac{b_{jk}b_{ki}}{b_{kk}} Z_j \vee \bigvee_{j \in \text{an}_{\text{nmw}}^U(i)} \bigvee_{k \in \text{De}(j) \cap \text{an}(i) \cap U} \frac{b_{jk}b_{ki}}{b_{kk}} Z_j \vee \bigvee_{j \in \text{An}_{\text{nmw}}^U(i)} b_{ji} Z_j \\ = \bigvee_{j \in \text{an}_{\text{mw}}^U(i)} b_{ji} Z_j \vee \bigvee_{j \in \text{an}_{\text{nmw}}^U(i)} \bigvee_{k \in \text{De}(j) \cap \text{an}(i) \cap U} \frac{b_{jk}b_{ki}}{b_{kk}} Z_j \vee \bigvee_{j \in \text{An}_{\text{nmw}}^U(i)} b_{ji} Z_j. \end{aligned} \quad (4.7)$$

For  $j \in \text{an}_{\text{nmw}}^U(i)$  we obtain by (2.7), as  $i \in U^c$ ,

$$\bigvee_{k \in \text{De}(j) \cap \text{an}(i) \cap U} \frac{b_{jk}b_{ki}}{b_{kk}} < b_{ji}. \quad (4.8)$$



Hence, we may omit the lower terms in (4.7) corresponding to the nodes of  $\text{an}_{\text{nmw}}^U(i)$ . This yields for the right-hand side of (4.6), since  $\text{An}_{\text{nmw}}^U(i) = \text{An}(i) \setminus \text{an}_{\text{nmw}}^U(i)$ ,

$$\bigvee_{j \in \text{an}_{\text{nmw}}^U(i)} b_{ji} Z_j \vee \bigvee_{j \in \text{An}_{\text{nmw}}^U(i)} b_{ji} Z_j = \bigvee_{j \in \text{An}(i)} b_{ji} Z_j = X_i.$$

It remains to show that (4.6) is the representation of  $X_i = \bigvee_{j \in \text{An}(i)} b_{ji} Z_j$  with minimal number of nodes in  $U$  and noise variables. We prove that we cannot omit any term in (4.6). Let  $j \in \text{an}_{\text{low}}^U(i)$ . Since  $\text{an}_{\text{low}}^U(i) \subseteq U$ , we know that  $j \in \text{an}_{\text{mw}}^U(i)$  and, hence,  $j \notin \text{An}_{\text{nmw}}^U(i)$ . Thus the term  $b_{ji} Z_j$  is only included in  $\bigvee_{k \in \text{an}_{\text{low}}^U(i)} \frac{b_{ki}}{b_{kk}} X_k$ . Assume that  $b_{ji} Z_j$  is only included in  $\frac{b_{ki}}{b_{kk}} X_k$  for some  $k \in \text{an}_{\text{low}}^U(i) \setminus \{j\}$ . This implies that  $\frac{b_{jk} b_{ki}}{b_{kk}} = b_{ji}$  and, hence, we have by (2.6),  $j \in \text{an}_{\text{mw}}^{\{k\}}(i)$ ; i.e., there exists a max-weighted path from  $j$  to  $i$  containing some node of  $U \setminus \{j, i\}$ . This is a contradiction to the fact that  $j \in \text{an}_{\text{low}}^U(i)$ . Thus  $b_{ji} Z_j$  is only included in  $\frac{b_{ji}}{b_{jj}} X_j$ , and we cannot omit this term in (4.6). Finally, observe from (4.7) and (4.8) that we also cannot omit any of the noise variables of  $\text{An}_{\text{nmw}}^U(i)$ .  $\square$

The following corollary presents the components of  $\mathbf{X}$  as minimal functions (with respect to the number of components) of their parent nodes.

**Corollary 4.8.** *Let  $\mathcal{D}^B$  be the minimal max-linear DAG from Theorem 3.5.F Then for  $i \in V$ ,  $\text{an}_{\text{low}}^{\text{pa}(i)}(i) = \text{pa}^B(i)$  and*

$$X_i = \bigvee_{k \in \text{pa}^B(i)} \frac{b_{ki}}{b_{kk}} X_k \vee b_{ii} Z_i = \bigvee_{k \in \text{pa}^B(i)} c_k^i X_k \vee c_i^i Z_i. \quad (4.9)$$

*Proof.* If  $k \in \text{pa}(i)$ , then either  $k \in \text{pa}^{\text{tr}}(i)$ ,  $k \in \text{pa}^B(i) \setminus \text{pa}^{\text{tr}}(i)$  or  $k \in \text{pa}(i) \setminus \text{pa}^B(i)$ . For  $k \in \text{pa}^{\text{tr}}(i)$  observe from the definition of the transitive reduction that  $[k \rightarrow i]$  is the only and, thus, max-weighted path from  $k$  to  $i$ . Hence,  $\text{pa}^{\text{tr}}(i) \subseteq \text{an}_{\text{low}}^{\text{pa}(i)}(i)$ . For  $k \in \text{pa}^B(i) \setminus \text{pa}^{\text{tr}}(i)$  we have by the definition of  $\mathcal{D}^B$  and (2.6) that  $k \notin \text{an}_{\text{mw}}^{\text{pa}(i) \setminus \{k\}}(i)$ . Thus the only max-weighted path from  $k$  to  $i$  is the edge  $[k \rightarrow i]$  and, therefore,  $k \in \text{an}_{\text{low}}^{\text{pa}(i)}(i)$ . For  $k \in \text{pa}(i) \setminus \text{pa}^B(i)$ , however, we know that  $k \in \text{an}_{\text{mw}}^{\text{pa}(i) \setminus \{k\}}(i)$ ; i.e., there exists a max-weighted path from  $k$  to  $i$  containing nodes of  $\text{pa}(i) \setminus \{i, j\}$  and, hence,  $k \in \text{pa}(i) \setminus \text{an}_{\text{low}}^{\text{pa}(i)}(i)$ . Altogether, we have  $\text{an}_{\text{low}}^{\text{pa}(i)}(i) = \text{pa}^B(i)$ .

Since every path from some  $j \in \text{an}(i)$  to  $i$  contains a node of  $\text{pa}(i)$ , it holds that  $\text{an}_{\text{mw}}^{\text{pa}(i)}(i) = \text{an}(i)$ ; i.e.,  $\text{An}_{\text{nmw}}^{\text{pa}(i)}(i) = \text{An}(i) \setminus \text{an}_{\text{mw}}^{\text{pa}(i)}(i) = \{i\}$ . Furthermore, recall from Theorem 3.5 that for  $k \in \text{pa}^B(i)$ ,  $c_k^i = \frac{b_{ki}}{b_{kk}}$ . Then, by Theorem 4.7, we obtain (4.9).  $\square$

**Remark 4.9.** Since (4.9) is the representation of  $X_i$  as a function of its parents  $\text{pa}(i)$  with minimal number of nodes in  $\text{pa}(i)$ , we recover again  $\mathcal{D}^B$  as the DAG with minimal number of edges corresponding to the SEM  $\mathbf{X}$ .  $\square$

The following example illustrates Theorem 4.7.

**Example 4.10.** [Continuation of Examples 2.1, 2.8 and 4.5: minimal representation by  $U$ ]

Consider  $U = \{1, 2\}$  and  $i = 4$ . Obviously, we have  $1, 2 \in \text{an}_{\text{mw}}^U(4)$ . Since there is no path from 3 to 4 which contains any node of  $U$ , we have  $\text{an}_{\text{mw}}^U(4) = \{1, 2\}$  and, hence,  $\text{An}_{\text{nmw}}^U(4) = \text{An}(i) \setminus \text{an}_{\text{mw}}^U(4) = \{3, 4\}$ . In Example 4.5 we have already determined the set  $\text{an}_{\text{low}}^U(4)$  depending on the ML coefficient matrix  $B = (b_{ij})_{4 \times 4}$ . Thus we obtain the two cases:

- (1) If  $b_{14} = \frac{b_{12}b_{24}}{b_{22}} \geq \frac{b_{13}b_{34}}{b_{33}}$ , then  $X_4 = \frac{b_{24}}{b_{22}} X_2 \vee b_{34} Z_3 \vee b_{44} Z_4$ .
- (2) If  $b_{14} = \frac{b_{13}b_{34}}{b_{33}} > \frac{b_{12}b_{24}}{b_{22}}$ , then  $X_4 = \frac{b_{24}}{b_{11}} X_1 \vee \frac{b_{24}}{b_{22}} X_2 \vee b_{34} Z_3 \vee b_{44} Z_4$ .

For  $U = \{2\}$  and  $i = 4$ , we have  $\text{an}_{\text{low}}^U(4) = \{2\}$ , since  $U \setminus \{2, 4\} = \emptyset$ . Obviously, we have  $2 \in \text{an}_{\text{mw}}^U(4)$ . Since there is no path from 3 to 4, which contains 2, we have that  $3, 4 \in \text{An}_{\text{nmw}}^U(4)$ . It remains to discuss node 1, which gives rise to the same two cases as above:

- (1) If  $b_{14} = \frac{b_{12}b_{24}}{b_{22}} \geq \frac{b_{13}b_{34}}{b_{33}}$ , then  $1 \in \text{an}_{\text{mw}}^U(4)$  and, consequently,  $\text{An}_{\text{nmw}}^U(i) = \{3, 4\}$ . Then  $X_4 = \frac{b_{24}}{b_{22}} X_2 \vee b_{34} Z_3 \vee b_{44} Z_4$ .
- (2) If  $b_{14} = \frac{b_{13}b_{34}}{b_{33}} > \frac{b_{12}b_{24}}{b_{22}}$ , then  $1 \notin \text{an}_{\text{mw}}^U(4)$  and, consequently,  $\text{An}_{\text{nmw}}^U(i) = \{1, 3, 4\}$ . Then  $X_4 = \frac{b_{24}}{b_{22}} X_2 \vee b_{14} Z_1 \vee b_{34} Z_3 \vee b_{44} Z_4$ .  $\square$

## 5. Order and ties between components of a max-linear graphical model

Throughout this section  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  is a ML graphical model with ML coefficient matrix  $B = (b_{ij})_{d \times d}$ . Denote again by  $(\Omega, \mathcal{A}, \mathbb{P})$  the probability space on which  $\mathbf{X}$  is defined.

Simply by definition of the ML graphical model the order between the weights in (1.3), the max-linear coefficients in (2.3), and also the components of  $\mathbf{X}$  plays an important role. From Lemma 4.1, for instance, we know that there is a natural order between the (appropriately scaled) components of  $\mathbf{X}$ , which holds on the whole of  $\Omega$ , and Theorem 2.11 shows conditions and properties for two random variables being equal on a subset of  $\Omega$ . In this section we take this a step further. For a given subset  $O \subseteq V$  we investigate the information which can be drawn from the fact that on a specific subset  $\Omega^o$  of  $\Omega$  some of the components of  $\mathbf{X}_O = (X_i, i \in O)$  may be equal, or one component may be strictly larger or smaller than another. We characterize this subset  $\Omega^o$  and recover on  $\Omega^o$  a reduced model, which may be easier to analyze. When we think of the nodes  $O$  as being observed in a statistical experiment, then this section provides probabilistic tools for predicting some part of the DAG which is not observed. Nodes in  $O$  naturally lead to information within all its ancestors  $\text{An}(O)$ , but even beyond certain conclusions are still possible.

### 5.1. Subgraphs of $\mathcal{D}$ induced by orders between nodes in $O \subseteq V$

We start by applying (4.2) to obtain for random variables in  $\text{An}(O)$  an upper bound based on the nodes in  $O$ :

$$X_j \leq \bigwedge_{i \in \text{De}(j) \cap O} \frac{b_{jj}}{b_{ji}} X_i =: X_j^u, \quad j \in \text{An}(O). \quad (5.1)$$

Note from (4.2) and the definition of  $X_j^u$  that

$$X_j^u = X_j \quad \text{for all } j \in O \quad \text{and} \quad X_j^u \leq \frac{b_{jj}}{b_{ji}} X_i \quad \text{for all } j \in \text{An}(O) \text{ and } i \in \text{De}(j) \cap O. \quad (5.2)$$

A similar concept has been used for conditional simulation of arbitrary max-linear models in [12]. This subsection is related to their important notion of *hitting scenarios*.

For  $j \in \text{An}(O)$  the upper bound  $X_j^u$  is not uniquely given by one node, but may be realized by several nodes. Obviously, the order between the components of  $\mathbf{X}_O$  has a direct influence on which random variables of  $O$  realize the upper bound  $X_j^u$ . Different realizations of  $\mathbf{X}_O$  may correspond to different upper bounds in (5.1) and in particular to different nodes in  $O$  attaining the bounds. We call  $\Omega^o \subseteq \Omega$  *order set* if for all  $j \in \text{An}(O)$  the upper bound  $X_j^u$  is realized by the same nodes of  $O$  on the whole of  $\Omega^o$ . The following provides a formal definition of an order set as well as the definition of an important subgraph corresponding to  $\Omega^o$ .

**Definition 5.1.** (a) Let  $\Omega^o \subseteq \Omega$ . If for all  $j \in \text{An}(O)$  and  $i \in \text{de}(j) \cap O$  it holds on  $\Omega^o$  that either  $X_j^u = \frac{b_{ji}}{b_{jj}} X_i$  or  $X_j^u < \frac{b_{ji}}{b_{jj}} X_i$ , and  $\Omega^o$  is the largest subset with this property, then we call  $\Omega^o$  *order set*.

(b) Let  $\Omega^o$  be some order set as defined in (a). Then the corresponding *order DAG*  $\mathcal{D}^o$  is the subgraph of  $\mathcal{D}$  with node set  $\text{An}(O)$  containing only those paths from  $\mathcal{D}$  which are max-weighted and have start and end node  $j \in \text{An}(O)$  and  $i \in O$ , respectively, satisfying  $X_j^u = \frac{b_{ji}}{b_{jj}} X_i$  on  $\Omega^o$ .  $\square$

Whereas the node set  $\text{An}(O)$  is given naturally by  $O$ , there is some freedom in choosing the edges of  $\mathcal{D}^o$ . Instead of keeping all paths from nodes  $j \in \text{An}(O)$  and  $i \in O$  satisfying  $X_j^u = \frac{b_{ji}}{b_{jj}} X_i$  on  $\Omega^o$ , we only keep the edges on max-weighted paths from  $j$  to  $i$ , since they result in a subgraph of  $\mathcal{D}$  containing only the relevant edges. This kind of construction will also allow us to find a simple characterization of the nodes which we can predict  $\mathbb{P}$ -a.s. (cf. Section 5.2).

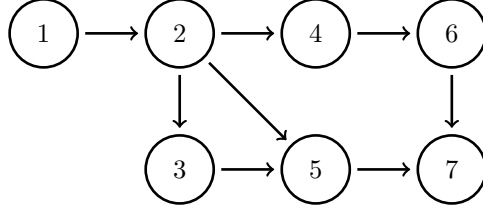
Note that, since the set  $V$  of nodes is finite, there exist only finitely many different order sets and, hence, also only finitely many different order DAGs.

**Remark 5.2.** Here and in what follows all node sets related to an order DAG  $\mathcal{D}^o$  are labelled by  $o$ , for instance, for  $i \in \text{An}(O)$  we denote by  $\text{an}^o(i)$ ,  $\text{pa}^o(i)$ , and  $\text{de}^o(i)$  the sets of the ancestors, parents, and descendants of  $i$  with respect to  $\mathcal{D}^o$  and, again,  $\text{An}^o(i) = \text{an}^o(i) \cup \{i\}$ . All other node sets related to some graphical concept, which are not labelled by  $o$ , are defined with respect to the original DAG  $\mathcal{D}$ .  $\square$

The following example illustrates the order DAGs implied by partial order information.

**Example 5.3.** Consider the ML graphical model  $(\mathcal{D}, \mathcal{L}(X_1, \dots, X_7))$  with DAG

$$\mathcal{D} = (\{1, 2, 3, 4, 5, 6, 7\}, \{(1, 2), (2, 3), (2, 4), (2, 5), (3, 5), (4, 6), (5, 7), (6, 7)\})$$



and ML coefficient matrix  $B = (b_{ij})_{7 \times 7}$ . Let  $O = \{4, 5\}$ ; hence,  $\text{An}(O) = \{1, 2, 3, 4, 5\}$ . Then the upper bounds in (5.1) are given by

$$X_1^u = \frac{b_{11}}{b_{14}}X_4 \wedge \frac{b_{11}}{b_{15}}X_5, \quad X_2^u = \frac{b_{22}}{b_{24}}X_4 \wedge \frac{b_{22}}{b_{25}}X_5, \quad X_3^u = \frac{b_{33}}{b_{35}}X_5, \quad X_4^u = X_4, \quad X_5^u = X_5.$$

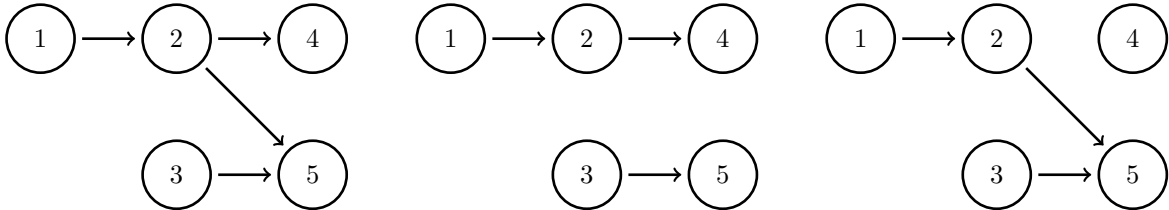
The three different scenarios  $X_1^u = \frac{b_{11}}{b_{14}}X_4 = \frac{b_{11}}{b_{15}}X_5$  or  $X_1^u = \frac{b_{11}}{b_{14}}X_4 < \frac{b_{11}}{b_{15}}X_5$  or  $X_1^u = \frac{b_{11}}{b_{15}}X_5 < \frac{b_{11}}{b_{14}}X_4$  yield the analogous situations for  $X_2^u$  and vice versa, which results in the three different order sets:

$$\begin{aligned} \Omega_1^o &= \{\omega \in \Omega : b_{15}X_4(\omega) = b_{14}X_5(\omega)\} = \{\omega \in \Omega : b_{25}X_4(\omega) = b_{24}X_5(\omega)\}, \\ \Omega_2^o &= \{\omega \in \Omega : b_{15}X_4(\omega) < b_{14}X_5(\omega)\} = \{\omega \in \Omega : b_{25}X_4(\omega) < b_{24}X_5(\omega)\}, \\ \Omega_3^o &= \{\omega \in \Omega : b_{15}X_4(\omega) > b_{14}X_5(\omega)\} = \{\omega \in \Omega : b_{25}X_4(\omega) > b_{24}X_5(\omega)\}. \end{aligned}$$

Next, we find the order DAGs from Definition 5.1(b) by identifying the max-weighted paths from  $j \in \text{An}(O)$  to all  $i \in \text{de}(j) \cap O$  which realize the bounds  $X_j^u$ . The edge  $[3 \rightarrow 5]$  is the only path from 3 to 5 and, hence, max-weighted. Observe that the sets of paths from 1 to 5 and from 2 to 5 are given by

$$P_{15} = \{[1 \rightarrow 2 \rightarrow 3 \rightarrow 5], [1 \rightarrow 2 \rightarrow 5]\} \quad \text{and} \quad P_{25} = \{[2 \rightarrow 3 \rightarrow 5], [2 \rightarrow 5]\}.$$

We assume that the DAG  $\mathcal{D}$  is minimal max-linear; i.e.,  $\mathcal{D} = \mathcal{D}^B$ . Since by Theorem 3.5,  $b_{25} > \frac{b_{23}b_{35}}{b_{33}}$ , the path  $[2 \rightarrow 3 \rightarrow 5]$  is not max-weighted, and the only max-weighted path from 2 to 5 is  $[2 \rightarrow 5]$ . Consequently, the only max-weighted path from 1 to 5 is  $[1 \rightarrow 2 \rightarrow 5]$ . Thus, corresponding to the realized bounds  $X_1^u, X_2^u$ , we obtain the following three possible order DAGs  $\mathcal{D}_1^o$ ,  $\mathcal{D}_2^o$ , and  $\mathcal{D}_3^o$  corresponding to the situations on  $\Omega_1^o$ ,  $\Omega_2^o$ , and  $\Omega_3^o$  from left to right:



$\square$

Some properties of an order DAG will be needed later on.

**Lemma 5.4.** Let  $\Omega^o$  be some order set with order DAG  $\mathcal{D}^o$ . Then

- (a) all paths in  $\mathcal{D}^o$  are max-weighted in  $\mathcal{D}$ , and
- (b) for  $i \in O$ ,  $j \in \text{An}^o(i)$  if and only if  $X_j^u = \frac{b_{jj}}{b_{ji}}X_i$  on  $\Omega^o$ .

*Proof.* (a) The statement follows directly from the construction of  $\mathcal{D}^o$  and Remark 2.7(iii).

(b) Note first from the definition of  $\mathcal{D}^o$  that, if  $X_j^u = \frac{b_{ji}}{b_{ji}} X_i$  on  $\Omega^o$ , then  $j \in \text{An}^o(i)$ .

In what follows all random variables are considered on  $\Omega^o$ . Assume there exists some  $j \in \text{An}^o(i)$  such that  $X_j^u \neq \frac{b_{ji}}{b_{ji}} X_i$ . By the definition of  $\mathcal{D}^o$ , however, there exists some  $k \in \text{An}(i)$  such that  $X_k^u = \frac{b_{kk}}{b_{ki}} X_i$  and there is a max-weighted path from  $k$  to  $i$  in  $\mathcal{D}$  which contains  $j$ ; i.e.,  $k \in \text{an}_{\text{mw}}^{\{j\}}(i)$ . Let  $l \in \text{De}(j) \cap O$ ; hence,  $l \in \text{De}(k) \cap O$ . Furthermore, observe from (2.6) and (2.7) that, since either  $k \in \text{an}_{\text{mw}}^{\{j\}}(l)$  or  $k \in \text{an}_{\text{nmw}}^{\{j\}}(l)$ ,  $b_{kl} \geq \frac{b_{kj}b_{jl}}{b_{jj}}$ . Thus altogether, also using (5.2) and (2.6), we obtain

$$\frac{b_{jj}}{b_{jl}} X_l \geq \frac{b_{kj}}{b_{kl}} X_l \geq \frac{b_{kj}}{b_{kk}} X_k^u = \frac{b_{kj}}{b_{ki}} X_i = \frac{b_{kj}b_{jj}}{b_{kj}b_{ji}} X_i = \frac{b_{jj}}{b_{ji}} X_i.$$

Consequently, for all  $l \in \text{De}(j) \cap O$ ,  $\frac{b_{jj}}{b_{jl}} X_l \leq \frac{b_{jj}}{b_{jl}} X_l$  and, hence,  $X_j^u = \bigwedge_{l \in \text{De}(j) \cap O} \frac{b_{jj}}{b_{jl}} X_l = \frac{b_{jj}}{b_{ji}} X_i$ , which is a contradiction to  $X_j^u \neq \frac{b_{jj}}{b_{ji}} X_i$ .  $\square$

**Definition 5.5.** Let  $\mathcal{D}^o$  be some order DAG. Define  $\mathbf{X}_{\text{An}(O)}^o$  as the random vector given by

$$X_i^o = \bigvee_{k \in \text{pa}^o(i)} c_k^i X_k^o \vee c_i^i Z_i, \quad i \in \text{An}(O),$$

with the same weights  $c_k^i$  and noise variables  $Z_i$  as in the representation (1.3) of  $\mathbf{X}$ . We call the resulting ML graphical model  $(\mathcal{D}^o, \mathcal{L}(\mathbf{X}_{\text{An}(O)}^o))$  *ML order graphical model*.  $\square$

Recall that  $\mathbf{X}_{\text{An}(O)}^o$  has a max-linear representation as in (2.3). The following is a consequence of Lemma 5.4(a) and Definition 2.6 and shows how the max-linear coefficients of  $\mathbf{X}$  are related to those of  $\mathbf{X}_{\text{An}(O)}^o$ .

**Lemma 5.6.** Let  $\Omega^o$  be some order set and  $(\mathcal{D}^o, \mathcal{L}(\mathbf{X}_{\text{An}(O)}^o))$  be the corresponding ML order graphical model. Denote by  $B^o = (b_{ij}^o)_{|\text{An}(O)| \times |\text{An}(O)|}$  the ML coefficient matrix of  $\mathbf{X}_{\text{An}(O)}^o$ . Then for all  $i \in \text{An}(O)$  and  $j \in \text{An}^o(i)$ ,  $b_{ji}^o = b_{ji}$ ; i.e.,

$$X_i^o = \bigvee_{j \in \text{An}^o(i)} b_{ji}^o Z_j = \bigvee_{j \in \text{An}^o(i)} b_{ji} Z_j, \quad i \in \text{An}(O). \quad (5.3)$$

It is remarkable that  $\mathbf{X}_O$  and  $\mathbf{X}_O^o$  coincide on  $\Omega^o$ .

**Proposition 5.7.** Let  $\Omega^o$  be some order set and  $(\mathcal{D}^o, \mathcal{L}(\mathbf{X}_{\text{An}(O)}^o))$  be the corresponding ML order graphical model. Then for  $i \in O$  we have on  $\Omega^o$  that  $X_i = X_i^o$ .

*Proof.* In what follows all random variables are considered on  $\Omega^o$ . Due to (5.3) it suffices to show that  $X_i = \bigvee_{j \in \text{An}(i)} b_{ji} Z_j = \bigvee_{j \in \text{An}^o(i)} b_{ji} Z_j$ . Let  $k \in \text{an}(i) \setminus \text{an}^o(i)$ . By Lemma 5.4(b) and (5.2) we have  $X_k^u < \frac{b_{kk}}{b_{ki}} X_i$ . Thus, also using (5.1), we obtain

$$\frac{b_{kk}}{b_{ki}} X_i > X_k^u \geq X_k = \bigvee_{j \in \text{An}(k)} b_{jk} Z_j \geq b_{kk} Z_k.$$

Consequently, we have for all  $k \in \text{an}(i) \setminus \text{an}^o(i)$ ,  $b_{ki} Z_k < X_i = \bigvee_{j \in \text{An}(i)} b_{ji} Z_j$  and, hence,  $X_i = \bigvee_{j \in \text{An}(i)} b_{ji} Z_j = \bigvee_{j \in \text{An}^o(i)} b_{ji} Z_j$ .  $\square$

We know from Theorem 2.2 that  $X_i = \bigvee_{j \in \text{An}(i)} b_{ji} Z_j$  holds for all  $i \in V$ . Proposition 5.7 implies that, on some order set  $\Omega^o$ , for all  $i \in O$  the maximum value of  $\{b_{ji} Z_j, j \in \text{An}(i)\}$  can only be achieved for some  $j \in \text{An}^o(i)$ ; i.e., we have for the node  $j(i)$  as in (2.8) that  $j(i) \in \text{An}^o(i)$ . Observe from the Definition 5.1(b) of  $\mathcal{D}^o$  and also from the proof of Proposition 5.7 that this model reduction originates in the upper bounds in (5.1). Now assume that there are two distinct nodes  $i_1, i_2 \in O$  such that  $\text{An}^o(i_1) \cap \text{An}^o(i_2) \neq \emptyset$ . So take some  $j \in \text{An}^o(i_1) \cap \text{An}^o(i_2)$ . By Lemma 5.4(b) we then have on  $\Omega^o$  that  $X_j^u = \frac{b_{jj}}{b_{ji_1}} X_{i_1} = \frac{b_{jj}}{b_{ji_2}} X_{i_2}$  and, consequently, by Theorem 2.11(b) that,  $\mathbb{P}$ -a.s. on  $\Omega^o$ ,  $j(i_1) = j(i_2) \in \text{An}^o(i_1) \cap \text{An}^o(i_2)$ . This shows that we may reduce the max-linear representation of  $\mathbf{X}_O$  even further. In the following we therefore investigate these upper bounds more closely and extend Theorem 2.11(b) to more than two nodes. To this end, we first introduce some graphical concepts.

**Definition 5.8.** Let  $\mathcal{D} = (V, E)$  be a DAG.

- (a) A DAG is called *weakly connected* if, replacing all of its directed edges with undirected edges, produces a connected graph; i.e., there is a (undirected) path between every pair of nodes. Moreover, a *weakly connected component* is a maximal weakly connected subgraph of a DAG.
- (b) Let  $A \subseteq V$ . A node in  $\bigcap_{i \in A} \text{An}(i)$  is a *common ancestor* of  $A$ , and we define  $\text{cAn}(A) := \bigcap_{i \in A} \text{An}(i)$ . A node  $j \in \text{cAn}(A)$  is a *lowest common ancestor* of  $A$  if there is no node  $l \in \text{cAn}(A)$  such that there exists a path from  $j$  to  $l$ . We denote the corresponding set by  $\text{lcAn}(A)$ .  $\square$

Assume that the order DAG  $\mathcal{D}^\circ$  has  $T \in \mathbb{N}$  weakly connected components, whose corresponding node sets we denote by  $C_1, \dots, C_T$ . They provide a partition of  $\text{An}(O)$  and  $O$  by

$$\text{An}(O) = \bigcup_{t=1}^T C_t \quad \text{and} \quad O = \bigcup_{t=1}^T (C_t \cap O) =: \bigcup_{t=1}^T O_t. \quad (5.4)$$

The following gives now a minimal max-linear representation of  $\mathbf{X}_O$  on  $\Omega^\circ$  with respect to the number of noise variables.

**Theorem 5.9.** Let  $\Omega^\circ$  be some order set and  $(\mathcal{D}^\circ, \mathcal{L}(\mathbf{X}_{\text{An}(O)}^\circ))$  the corresponding ML order graphical model. Assume that the weakly connected components of  $\mathcal{D}^\circ$  have node sets  $C_1, \dots, C_T$ , and recall (5.4). Then the following holds  $\mathbb{P}$ -a.s. on  $\Omega^\circ$  for every  $t \in \{1, \dots, T\}$ :

- (a) For all  $i \in O_t$  the node  $j(i)$  as in (2.8) is the same unique node and belongs to  $\text{cAn}^\circ(O_t)$ .
- (b) For every  $i \in O_t$  we have

$$X_i = \bigvee_{j \in \text{cAn}^\circ(O_t)} b_{ji} Z_j = \bigvee_{k \in \text{lcAn}^\circ(O_t)} \frac{b_{ki}}{b_{kk}} X_k^\circ. \quad (5.5)$$

- (c) The number of noise variables  $Z_j$  in representation (5.5) is minimal. In particular, the maximum occurs in every  $Z_j$  for  $j \in \text{cAn}^\circ(O_t)$  with positive probability.

*Proof.* Assume wlog that  $\mathbb{P}(\Omega^\circ) > 0$ . Restrict all random variables to  $\Omega^\circ$ , and let  $t \in \{1, \dots, T\}$  be fixed. First, suppose that  $|O_t| = 1$ ; i.e.,  $O_t = \{i\}$  for some  $i \in V$ . Observe from Definition 5.8(b) that  $\text{cAn}^\circ(O_t) = \text{cAn}^\circ(\{i\}) = \text{An}^\circ(i)$  and, since for every  $j \in \text{An}^\circ(i)$  there exists a path from  $j$  to  $i$  in  $\mathcal{D}^\circ$ , that  $\text{lcAn}^\circ(O_t) = \text{lcAn}^\circ(\{i\}) = \{i\}$ . Thus (5.5) follows directly from (5.3) and Proposition 5.7. From (5.5) and Theorem 2.11(a) we get (a).

Now suppose that  $|O_t| \geq 2$ , and let  $i_1, i_2 \in O_t$  and  $i_1 \neq i_2$ . By Definition 5.1(b), as  $i_1$  and  $i_2$  are in the same weakly connected component of  $\mathcal{D}^\circ$ , there must exist nodes

$$i_1 = l_0, l_1, \dots, l_m = i_2 \in O_t \quad \text{and} \quad j_s \in \text{An}^\circ(l_{s-1}) \cap \text{An}^\circ(l_s), \quad s = 1, \dots, m,$$

such that

$$X_{j_s}^u = \frac{b_{j_s j_s}}{b_{j_s l_{s-1}}} X_{l_{s-1}} = \frac{b_{j_s j_s}}{b_{j_s l_s}} X_{l_s}, \quad s = 1, \dots, m.$$

First, assume that  $m = 1$ ; i.e.,

$$X_j^u = \frac{b_{jj}}{b_{ji_1}} X_{i_1} = \frac{b_{jj}}{b_{ji_2}} X_{i_2} \quad \text{for some } j \in \text{An}^\circ(i_1) \cap \text{An}^\circ(i_2).$$

By Theorem 2.11(b) we have  $\mathbb{P}$ -a.s. on  $\Omega^\circ$  that  $j(i_1) = j(i_2)$ . Furthermore, we know from (5.3) and Proposition 5.7 that  $j(i_1) \in \text{An}^\circ(i_1)$  and  $j(i_2) \in \text{An}^\circ(i_2)$ . Hence,  $\mathbb{P}$ -a.s. on  $\Omega^\circ$ ,  $j(i_1) = j(i_2) \in \text{cAn}^\circ(\{i_1, i_2\}) = \text{An}^\circ(i_1) \cap \text{An}^\circ(i_2)$ .

For  $m > 1$  we may conclude from above that,  $\mathbb{P}$ -a.s. on  $\Omega^\circ$ ,

$$j(i_1) = j(l_0) = j(l_1) = \dots = j(l_m) = j(i_2) \in \bigcap_{s \in \{1, \dots, m\}} \text{An}^\circ(l_s),$$

since any finite union of null sets is a null set.

This argument applies to all different pairs of nodes in  $O_t$  such that  $j(i)$  is the same node for all  $i \in O_t$ , and, hence, by (5.3) and Proposition 5.7,  $j(i) \in \bigcap_{i \in O_t} \text{An}^o(i) = \text{cAn}^o(O_t)$ . Its uniqueness is already given in Theorem 2.11(a). Hence, we have proved (a).

The first equality of (5.5) follows from (a), and also (c) follows immediately from the above construction.

Thus it remains to show that the second equality of (5.5) holds. Observe from the Definition 5.8(b) of the lowest common ancestors that

$$\text{cAn}^o(O_t) = \bigcup_{k \in \text{lcAn}^o(O_t)} \text{An}^o(k).$$

We therefore obtain for  $i \in O_t$  that,  $\mathbb{P}$ -a.s. on  $\Omega^o$ ,

$$X_i = \bigvee_{j \in \text{cAn}^o(O_t)} b_{ji} Z_j = \bigvee_{k \in \text{lcAn}^o(O_t)} \bigvee_{j \in \text{An}^o(k)} b_{ji} Z_j.$$

Now let  $k \in \text{lcAn}^o(O_t)$  and  $j \in \text{An}^o(k)$ . Observe from Lemma 5.4(a) that in  $\mathcal{D}$  there exists a max-weighted path from  $j$  to  $i$  containing  $k$ ; hence,  $j \in \text{an}_{\text{mw}}^{\{k\}}(i)$ . Thus by (2.6) and (5.3) we have  $\mathbb{P}$ -a.s. on  $\Omega^o$ ,

$$X_i = \bigvee_{k \in \text{lcAn}^o(O_t)} \left( \bigvee_{j \in \text{An}^o(k)} \frac{b_{jk} b_{ki}}{b_{kk}} Z_j \vee b_{kk} Z_k \right) = \bigvee_{k \in \text{lcAn}^o(O_t)} \frac{b_{ki}}{b_{kk}} \left( \bigvee_{j \in \text{An}^o(k)} b_{jk} Z_j \right) = \bigvee_{k \in \text{lcAn}^o(O_t)} \frac{b_{ki}}{b_{kk}} X_k^o.$$

□

## 5.2. Almost sure prediction

Let  $\Omega^o$  be some order set with order DAG  $\mathcal{D}^o$ . From Theorem 5.9(a) we know that for every node  $i \in O$  the maximum value over all  $\{b_{ji} Z_j, j \in \text{An}(i)\}$  can  $\mathbb{P}$ -a.s. on  $\Omega^o$  only be achieved for those  $j$  which are in  $\mathcal{D}^o$  common ancestors of all nodes of  $O$  contained in the same weakly connected component; i.e.,  $j(i) \in \text{cAn}^o(O_t)$ . In the following, we investigate the information which can be drawn from this fact for nodes outside  $O$ , i.e. for nodes of  $O^c = V \setminus O$ . As a motivation we present an example.

**Example 5.10.** Consider the ML graphical model  $(\mathcal{D}, \mathcal{L}(X_1, X_2, X_3, X_4))$  with DAG



and ML coefficient matrix  $B = (b_{ij})_{4 \times 4}$ . Let  $O = \{2, 4\}$ . Find out that the order DAG corresponding to the order set  $\Omega^o = \{\omega \in \Omega : b_{14} X_2(\omega) = b_{12} X_4(\omega)\}$  is given by  $\mathcal{D}^o = \mathcal{D}$ . The following holds  $\mathbb{P}$ -a.s. on  $\Omega^o$ . Theorem 5.9(b) yields that

$$X_4 = \bigvee_{k \in \text{lcAn}^o(\{2, 4\})} \frac{b_{ki}}{b_{kk}} X_k = \frac{b_{14}}{b_{11}} X_1.$$

Thus, using (3.5), we obtain  $\frac{b_{33}}{b_{34}} X_4 = \frac{b_{13}}{b_{11}} X_1$ . Furthermore, we know from (4.2) that

$$\frac{b_{13}}{b_{11}} X_1 \leq X_3 \leq \frac{b_{33}}{b_{34}} X_4.$$

Since the lower and upper bounds are equal,  $X_3 = \frac{b_{33}}{b_{34}} X_4$  must hold. Altogether, we obtain

$$X_1 = \frac{b_{11}}{b_{14}} X_4 = \frac{b_{11}}{b_{12}} X_2 \quad \text{and} \quad X_3 = \frac{b_{33}}{b_{34}} X_4.$$

We denote this extension of  $O$  by  $\overline{O} = \{1, 2, 3, 4\}$ .

□

We formulate a general result on nodes outside  $O$  leading to this extension of  $O$ .

**Theorem 5.11.** *Assume the same conditions as in Theorem 5.9 hold. For some  $t \in \{1, \dots, T\}$  let  $l \in C_t \setminus O$  such that  $\text{cAn}^o(O_t) \subseteq \text{An}^o(l)$ . Then the following holds  $\mathbb{P}$ -a.s. on  $\Omega^o$ :*

- (a) *The node  $j(l)$  as in (2.8) is unique and  $j(l) = j(i)$  for all  $i \in O_t$ . Moreover,  $j(l) \in \text{cAn}^o(O_t)$ .*
- (b) *We have*

$$X_l = X_l^o = X_l^u = \bigvee_{j \in \text{cAn}^o(O_t)} b_{jl} Z_j = \bigvee_{k \in \text{lcAn}^o(O_t)} \frac{b_{kl}}{b_{kk}} X_k^o. \quad (5.6)$$

*Proof.* Assume wlog that  $\mathbb{P}(\Omega^o) > 0$ . The following holds  $\mathbb{P}$ -a.s. on  $\Omega^o$ . From Definition 5.1(b) we know that all terminal nodes of  $\mathcal{D}^o$  belong to  $O$ . Thus, we have that  $\text{de}^o(l) \cap O_t \neq \emptyset$ . So take some  $i \in \text{de}^o(l) \cap O_t$ . From Lemma 5.4(b), as  $l \in \text{An}^o(i)$ , and (5.5) we obtain

$$X_l^u = \frac{b_{li}}{b_{li}} X_i = \bigvee_{j \in \text{cAn}^o(O_t)} \frac{b_{li} b_{ji}}{b_{li}} Z_j = \bigvee_{k \in \text{lcAn}^o(O_t)} \frac{b_{li} b_{ki}}{b_{li} b_{kk}} X_k^o.$$

Let  $j \in \text{cAn}^o(O_t)$ . Since  $\text{cAn}^o(O_t) \subseteq \text{An}^o(l)$ , there exists a path in  $\mathcal{D}^o$  from  $j$  to  $i$  containing node  $l$ . Hence, by Lemma 5.4(a) there is also a max-weighted path in  $\mathcal{D}$  from  $j$  to  $i$  containing  $l$ ; i.e.,  $j \in \text{an}_{\text{mw}}^{\{l\}}(i)$ . Thus, as  $\text{lcAn}^o(O_t) \subseteq \text{cAn}^o(O_t)$ , we obtain with (2.6),

$$X_l^u = \bigvee_{j \in \text{cAn}^o(O_t)} b_{jl} Z_j = \bigvee_{k \in \text{lcAn}^o(O_t)} \frac{b_{kl}}{b_{kk}} X_k^o.$$

Furthermore, using first (5.1), then  $\text{An}^o(l) \subseteq \text{An}(l)$  as well as (5.3), and finally  $\text{cAn}^o(O_t) \subseteq \text{An}^o(l)$  yields

$$X_l^u \geq X_l = \bigvee_{j \in \text{An}(l)} b_{jl} Z_j \geq \bigvee_{j \in \text{An}^o(l)} b_{jl} Z_j = X_l^o \geq \bigvee_{j \in \text{cAn}^o(O_t)} b_{jl} Z_j = X_l^u.$$

Since the lower and upper bounds are equal, we have shown (5.6). From above we also see that  $X_l = \frac{b_{li}}{b_{li}} X_i$ , which implies by Theorem 2.11(b) that  $j(l) = j(i)$ . Then by Theorem 5.9(a) we have (a).  $\square$

Define the set

$$\tilde{O} := \{l \in V : X_l = aX_i \text{ } \mathbb{P}\text{-a.s. on } \Omega^o \text{ for some } a \in \mathbb{R}_+ \text{ and } i \in O\}.$$

When we think of the nodes  $O$  as being observed in a statistical experiment, then  $\tilde{O}$  contains exactly those nodes which we can predict, at least  $\mathbb{P}$ -a.s. The following shows that  $\tilde{O}$  contains exactly the nodes from  $O$  and Theorem 5.11.

**Theorem 5.12.** *Assume the same conditions as in Theorem 5.9 hold. Define the extended  $O$ -set*

$$\overline{O} := \bigcup_{t=1, \dots, T} \{l \in C_t : \text{cAn}^o(O_t) \subseteq \text{An}^o(l)\}. \quad (5.7)$$

Then

$$\overline{O} = \tilde{O} \subseteq \text{An}(O). \quad (5.8)$$

Moreover, the following assertions hold  $\mathbb{P}$ -a.s. on  $\Omega^o$  for every  $t \in \{1, \dots, T\}$ :

- (a) *For all  $i \in \overline{O}_t := C_t \cap \overline{O}$  the node  $j(i)$  as in (2.8) is the same unique node and belongs to  $\text{cAn}^o(O_t)$ .*
- (b) *We have for  $i \in \overline{O}_t$ ,*

$$X_i = X_i^o = X_i^u = \bigvee_{j \in \text{cAn}^o(O_t)} b_{ji} Z_j = \bigvee_{k \in \text{lcAn}^o(O_t)} \frac{b_{ki}}{b_{kk}} X_k^o. \quad (5.9)$$

*Proof.* Assume wlog that  $\mathbb{P}(\Omega^o) > 0$ . First, observe that all nodes of  $O$  are obtained in the sets  $\tilde{O}$  and  $\overline{O}$ . Thus it remains to show that  $\overline{O} \setminus O = \tilde{O} \setminus O$ .



Let  $l \in \overline{O} \setminus O$ ; i.e.,  $l \in C_t \setminus O$  for some  $t \in \{1, \dots, T\}$  and  $\text{cAn}^o(O_t) \subseteq \text{An}^o(l)$ . We know from (5.6) and the definition of  $X_l^u$  in (5.1) that  $\mathbb{P}$ -a.s. on  $\Omega^o$ ,

$$X_l = X_l^u = \bigwedge_{i \in \text{De}(l) \cap O} \frac{b_{li}}{b_{li}} X_i.$$

Hence,  $l \in \widetilde{O} \setminus O$ .

Let now  $l \in (\overline{O} \setminus O)^c$ . To show that  $l \in (\widetilde{O} \setminus O)^c$ , by Theorem 2.11(b) it suffices to construct an event  $E \subseteq \Omega^o$  with positive probability such that on  $E$  we have  $j(l) \neq j(i)$  for all  $i \in O$ . We consider the cases  $l \in \text{An}(O)$  and  $l \notin \text{An}(O)$  separately.

Let  $l \in \text{An}(O)$ ; i.e.,  $l \in C_{t_l}$  for some  $t_l \in \{1, \dots, T\}$ . As  $l \in (\overline{O} \setminus O)^c$ , we also have  $\text{cAn}^o(O_{t_l}) \not\subseteq \text{An}^o(l)$ . Recall from Theorem 5.9(a) that for all  $t \in \{1, \dots, T\}$  the set  $\text{cAn}^o(O_t)$  is non-empty; i.e., there exists some  $j_t \in \text{cAn}^o(O_t)$ . For  $t \in \{1, \dots, T\} \setminus \{t_l\}$  let  $j_t \in \text{cAn}^o(O_t) \subseteq C_t$  be some arbitrary but fixed node, and let  $j_{t_l} \in \text{cAn}^o(O_{t_l}) \setminus \text{An}^o(l)$ , which exists as  $\text{cAn}^o(O_{t_l}) \not\subseteq \text{An}^o(l)$ . Since the sets  $C_1, \dots, C_T$  are disjoint and  $j_{t_l} \notin \text{An}^o(l)$ , we have  $j_t \neq l$  for all  $t \in \{1, \dots, T\}$ . Observe from Theorem 5.9 that

$$E_1 = \{\omega \in \Omega : b_{j_t i} Z_{j_t}(\omega) > \bigvee_{j \in \text{An}(i) \setminus \{j_t\}} b_{ji} Z_j(\omega) \forall t \in \{1, \dots, T\} \text{ and } i \in O_t\}$$

with  $\mathbb{P}(E_1) > 0$ . Immediately by construction we have  $E_1 \subseteq \Omega^o$ .

Define  $E_2 = \{\omega \in \Omega : b_{li} Z_l(\omega) > \bigvee_{j \in \text{An}(l)} b_{jl} Z_j(\omega)\}$ . On  $E_1 \cap E_2$  we have  $j(l) = l \neq j(i) = j_t$  for all  $t \in \{1, \dots, T\}$  and  $i \in O_t$ ; i.e.,  $j(l) \neq j(i)$  for all  $i \in O$ . Thus it remains to show that  $\mathbb{P}(E_1 \cap E_2) > 0$ . Since the noise variables are continuous with support  $\mathbb{R}_+$ ,  $\mathbb{P}(E = E_1 \cap E_2) = 0$  could happen if and only if  $E_1$  and  $E_2$  are contradictory. From the definitions of  $E_1$  and  $E_2$  this would only be the case if and only if there exist some  $t \in \{1, \dots, T\}$  and  $i \in O_t$  such that  $l \in \text{An}(i) \setminus \{j_t\}$ ,  $j_t \in \text{an}(l)$ , and  $b_{ji} = \frac{b_{j_t l} b_{li}}{b_{li}}$ . This would lead to a contradiction for this constellation of nodes due to the fact that  $E_1$  and  $E_2$  imply the following two inequalities, respectively,

$$b_{j_t i} Z_{j_t} > b_{li} Z_l > \frac{b_{j_t l} b_{li}}{b_{li}} Z_{j_t},$$

which indeed contradicts  $b_{ji} = \frac{b_{j_t l} b_{li}}{b_{li}}$ . Thus it remains to show that for all  $t \in \{1, \dots, T\}$  and  $i \in O_t$  such that  $j_t \in \text{an}(l)$  and  $l \in \text{An}(i) \setminus \{j_t\}$  we have  $b_{ji} > \frac{b_{j_t l} b_{li}}{b_{li}}$  or, equivalently, by (2.7)  $j_t \in \text{an}_{\text{nmw}}^{\{l\}}(i)$ . We consider the nodes  $i \in O_{t_l}$  with this property separately. First, let  $t \in \{1, \dots, T\} \setminus \{t_l\}$ , and assume that there exists some  $i \in O_t$  such that  $j_t \in \text{an}(l)$  and  $l \in \text{An}(i) \setminus \{j_t\}$ . Hence, there exists at least one path  $p$  from  $j$  to  $i$  containing node  $l$ . Assume that  $p$  is max-weighted. As  $j_t \in \text{cAn}^o(O_t)$ , we have by Lemma 5.4(b)  $X_{j_t}^u = \frac{b_{j_t j_t}}{b_{j_t i}} X_i$ , and, hence, by the construction of  $\mathcal{D}^o$  (cf. Definition 5.1(b)) that  $l \in C_t$ . This is, however, a contradiction to the fact that  $l \in C_{t_l}$  and, hence,  $p$  is not max-weighted; i.e.,  $j_t \in \text{an}_{\text{nmw}}^{\{l\}}(i)$ .

Assume now that there exists some  $i \in O_{t_l}$  such that  $j_{t_l} \in \text{an}(l)$  and  $l \in \text{An}(i) \setminus \{j_{t_l}\}$ . Observe again from Definition 5.1(b), as  $j_{t_l} \in \text{an}(l)$  but  $j_{t_l} \notin \text{An}^o(l)$ , that there exists no max-weighted path from  $j_{t_l}$  to  $i$  containing  $l$ ; i.e.,  $j_{t_l} \in \text{an}_{\text{nmw}}^{\{l\}}(i)$ . Thus we have shown that for all  $l \in (\overline{O} \setminus O)^c \cap \text{An}(O)$ ,  $l \in (\widetilde{O} \setminus O)^c$ .

Let now  $l \in (\overline{O} \setminus O)^c$  with  $l \notin \text{An}(O)$ . We may proceed similarly as in the previous case. For  $t \in \{1, \dots, T\}$  let  $j_t \in \text{cAn}^o(O_t)$  be arbitrary but fixed nodes. Since  $\bigcup_{t=1}^T \text{cAn}^o(O_t) \subseteq \text{An}(O)$ ,  $j_t \neq l$  for all  $t \in \{1, \dots, T\}$ . Define the sets  $E_1$  and  $E_2$  as above, and observe that  $E_1$  and  $E_2$  are not contradictory, since the existence of  $t \in \{1, \dots, T\}$  and  $i \in O_t$  such that  $j_t \in \text{an}(l)$  and  $l \in \text{An}(i) \setminus \{j_t\}$  is a contradiction to the fact that  $l \notin \text{An}(O)$ .

Thus we have shown that  $\overline{O} = \widetilde{O}$ . Since by (5.4)  $\bigcup_{t=1}^T C_t = \text{An}(O)$ , we have  $\overline{O} \subseteq \text{An}(O)$  and, hence, (5.8).

We have already shown the assertions (a) and (b) in (5.2), Proposition 5.7, and Theorems 5.9 and 5.11 (depending on whether  $i \in O$  or  $i \in \overline{O} \setminus O$ ).  $\square$

The following provides a representation of the nodes of  $\overline{O}$  contained in the same weakly connected component, which is needed later on.

**Proposition 5.13.** *Assume the same conditions as in Theorem 5.9 hold, and recall (5.7). For  $t = 1, \dots, T$  let  $i_t \in O_t$  and  $j_t \in \text{cAn}^o(O_t)$  be arbitrary but fixed nodes. Then  $\mathbb{P}$ -a.s. on  $\Omega^o$  we have for all  $t \in \{1, \dots, T\}$  and  $i \in \overline{O}_t = C_t \cap \overline{O}$ ,*

$$X_i = X_i^u = \frac{b_{j_t i}}{b_{j_t i_t}} X_{i_t} = \frac{b_{j_t i}}{b_{j_t j_t}} X_{j_t}^u. \quad (5.10)$$

*Proof.* Assume wlog that  $\mathbb{P}(\Omega^o) > 0$ . The following holds  $\mathbb{P}$ -a.s. on  $\Omega^o$ . Let  $t \in \{1, \dots, T\}$  be fixed. The first equality of (5.10) is already given in (5.9). Recall from Theorem 5.9(a) that  $\text{cAn}^o(O_t)$  is non-empty; i.e., there exists some  $j_t \in \text{cAn}^o(O_t)$ . Furthermore, observe that  $O_t \neq \emptyset$ , since all terminal nodes of  $\mathcal{D}^o$  belong to  $O$ ; i.e., there exists some  $i_t \in O_t$ .

Let  $i \in O_t$ . By Lemma 5.4(b), as  $j_t \in \text{An}^o(i) \cap \text{An}^o(i_t)$ , we have  $X_{j_t}^u = \frac{b_{j_t j_t}}{b_{j_t i}} X_i = \frac{b_{j_t j_t}}{b_{j_t i_t}} X_{i_t}$ , which implies that  $X_i = \frac{b_{j_t i}}{b_{j_t i_t}} X_{i_t} = \frac{b_{j_t i}}{b_{j_t j_t}} X_{j_t}^u$ .

Now let  $i \in \overline{O}_t \setminus O_t$ . Since all terminal nodes of  $\mathcal{D}^o$  belong to  $O$ , there exists some  $k \in \text{de}^o(i) \cap O_t$ . Thus from Lemma 5.4(b), as  $i \in \text{An}^o(k)$ , and (5.10), which we have already shown for  $k \in O_t$ , we obtain

$$X_i^u = \frac{b_{ii}}{b_{ik}} X_k = \frac{b_{ii} b_{j_t k}}{b_{ik} b_{j_t i_t}} X_{i_t} = \frac{b_{ii} b_{j_t k}}{b_{ik} b_{j_t j_t}} X_{j_t}^u.$$

Now observe from Lemma 5.4(a) that, as by (5.7)  $j_t \in \text{An}^o(i)$ , there exists a max-weighted path in  $\mathcal{D}$  from  $j_t$  to  $k$  containing  $i$ ; i.e.,  $j_t \in \text{an}_{\text{mw}}^{\{i\}}(k)$ . Thus by (2.6) we have

$$X_i^u = \frac{b_{ii}}{b_{ik}} X_k = \frac{b_{j_t i}}{b_{j_t i_t}} X_{i_t} = \frac{b_{j_t i}}{b_{j_t j_t}} X_{j_t}^u.$$

□

### 5.3. What we know about nodes outside $\overline{O}$

Let  $\Omega^o$  be some order set with order DAG  $\mathcal{D}^o$ . From Theorem 5.12 we know that the extended  $O$ -set  $\overline{O}$  contains exactly those nodes which are at least  $\mathbb{P}$ -a.s. on  $\Omega^o$  equal to some appropriately scaled node of  $O$ . Next, we investigate what information we can still deduce for nodes outside  $\overline{O}$  from nodes in it.

Assume that  $\overline{O} \subsetneq V$ . Then by setting  $U = \overline{O}$ , Theorem 4.7 provides a representation for the nodes outside the set  $\overline{O}$ , which holds  $\mathbb{P}$ -a.s. on  $\Omega^o$ , as a function of the nodes of  $\overline{O}$  and some noise variables. This allows for a substantial reduction in the representation of certain random variables, however, on some notational costs.

**Corollary 5.14.** *Assume the same conditions as in Proposition 5.13 hold. Recall Definition 4.3(a). Then  $\mathbb{P}$ -a.s. on  $\Omega^o$  we have for  $i \in \overline{O}^c$ ,*

$$\begin{aligned} X_i &= \bigvee_{k \in \text{an}_{\text{low}}^{\overline{O}}(i)} \frac{b_{ki}}{b_{kk}} X_k^u \vee \bigvee_{j \in \text{An}_{\text{nmw}}^{\overline{O}}(i)} b_{ji} Z_j \\ &= \bigvee_{t=1, \dots, T} \left( \bigvee_{k \in \text{an}_{\text{low}}^{\overline{O}}(i) \cap C_t} \frac{b_{j_t k} b_{ki}}{b_{j_t j_t} b_{kk}} \right) X_{j_t}^u \vee \bigvee_{j \in \text{An}_{\text{nmw}}^{\overline{O}}(i)} b_{ji} Z_j \\ &= \bigvee_{t=1, \dots, T} \left( \bigvee_{k \in \text{an}_{\text{low}}^{\overline{O}}(i) \cap C_t} \frac{b_{j_t k} b_{ki}}{b_{j_t i_t} b_{kk}} \right) X_{i_t} \vee \bigvee_{j \in \text{An}_{\text{nmw}}^{\overline{O}}(i)} b_{ji} Z_j. \end{aligned}$$

*Proof.* Observe from (5.7) that every  $k \in \text{an}_{\text{low}}^{\overline{O}}(i) \subseteq \overline{O}$  belongs to some  $C_t$ . Thus we obtain by (4.6),

$$X_i = \bigvee_{k \in \text{an}_{\text{low}}^{\overline{O}}(i)} \frac{b_{ki}}{b_{kk}} X_k \vee \bigvee_{j \in \text{An}_{\text{nmw}}^{\overline{O}}(i)} b_{ji} Z_j = \bigvee_{t=1, \dots, T} \bigvee_{k \in \text{an}_{\text{low}}^{\overline{O}}(i) \cap C_t} \frac{b_{ki}}{b_{kk}} X_k \vee \bigvee_{j \in \text{An}_{\text{nmw}}^{\overline{O}}(i)} b_{ji} Z_j. \quad (5.11)$$

Then by (5.10) we have  $\mathbb{P}$ -a.s. on  $\Omega^o$  the three representations. □

We proceed with two examples, which indicate that there is still some room for reduction in our representations.

**Example 5.15.** [Continuation of Examples 2.1, 2.8, 4.5, 4.10: a new representation]

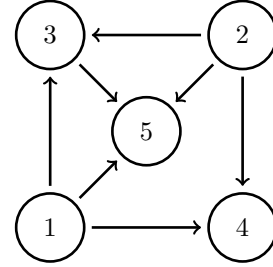
Assume that  $O = \{2, 3\}$  and  $\Omega^o = \{\omega \in \Omega : b_{13}X_2(\omega) = b_{12}X_3(\omega)\}$ . Hence,  $\mathcal{D}^o = (\{1, 2, 3\}, \{(1, 2), (1, 3)\})$ . By Corollary 5.14 for  $i = 4$  we obtain  $\mathbb{P}$ -a.s. on  $\Omega^o$ ,

$$X_4 = \left( \frac{b_{12}b_{24}}{b_{11}b_{22}} \vee \frac{b_{13}b_{34}}{b_{11}b_{33}} \right) X_1^u \vee b_{44}Z_4 = \frac{b_{14}}{b_{11}} X_1^u \vee b_{44}Z_4,$$

where we have applied (3.5) for the last equality.  $\square$

**Example 5.16.** Consider the ML graphical model  $(\mathcal{D}, \mathcal{L}(X_1, X_2, X_3, X_4, X_5))$  with DAG

$$\mathcal{D} = (\{1, 2, 3, 4, 5\}, \{(1, 3), (1, 4), (1, 5), (2, 3), (2, 4), (3, 5)\})$$



and ML coefficient matrix  $B = (b_{ij})_{4 \times 4}$ . Let  $O = \{3, 4\}$ . Consider the order set  $\Omega^o = \{b_{14}X_3 = b_{13}X_4 \text{ and } b_{24}X_3 = b_{23}X_4\}$ , then we find

$$\mathcal{D}^o = (\{1, 2, 3, 4\}, \{(1, 3), (1, 4), (2, 3), (2, 4)\}).$$

We assume that the DAG  $\mathcal{D}$  is minimal max-linear; i.e.,  $\mathcal{D} = \mathcal{D}^B$  (cf. Theorem 3.5). Then, as  $\text{an}_{\text{nmw}}^{\overline{O}}(5) = \{1, 2\}$ , we obtain by Theorem 4.7 for  $i = 5$  that  $X_5 = \frac{b_{35}}{b_{33}} X_3 \vee b_{15}Z_1 \vee b_{25}Z_2 \vee b_{55}Z_5$ . Furthermore, we have by (5.5)  $\mathbb{P}$ -a.s. on  $\Omega^o$ ,  $X_3 = b_{13}Z_1 \vee b_{23}Z_2$ . Finally, since by Theorem 3.5  $b_{15} > \frac{b_{13}b_{35}}{b_{33}}$  and  $b_{25} > \frac{b_{23}b_{35}}{b_{33}}$ , we obtain  $\mathbb{P}$ -a.s. on  $\Omega^o$ ,

$$X_5 = \frac{b_{35}}{b_{33}} (b_{13}Z_1 \vee b_{23}Z_2) \vee b_{15}Z_1 \vee b_{25}Z_2 \vee b_{55}Z_5 = b_{15}Z_1 \vee b_{25}Z_2 \vee b_{55}Z_5.$$

Examples 5.15 and 5.16 indicate that the decomposition of Corollary 5.14 can be further reduced. The following is our final result in this direction.  $\square$

**Proposition 5.17.** Assume the same conditions as in Theorem 5.9 hold, and recall (5.7). For  $i \in \overline{O}^c$  and  $t = 1, \dots, T$  let  $j_{t,i}^{\text{mw}}$  be a common ancestor of  $O_t$ , which is at the same time an ancestor of  $i$ , which is max-weighted by  $\overline{O}_t$ , i.e.  $j_{t,i}^{\text{mw}} \in \text{an}_{\text{mw}}^{\overline{O}_t}(i) \cap \text{cAn}^o(O_t)$ , and  $i_t \in O_t$ , both arbitrary but fixed. Then we have  $\mathbb{P}$ -a.s. on  $\Omega^o$ ,

$$\begin{aligned} X_i &= \bigvee_{\substack{t=1, \dots, T: \\ (\text{an}_{\text{mw}}^{\overline{O}_t}(i) \cap \text{cAn}^o(O_t)) \neq \emptyset}} \frac{b_{j_{t,i}^{\text{mw}} i}}{b_{j_{t,i}^{\text{mw}} i_t} j_{t,i}^{\text{mw}}} X_{j_{t,i}^{\text{mw}}}^u \vee \bigvee_{j \in \text{An}_{\text{nmw}}^{\overline{O}}(i)} b_{ji} Z_j \\ &= \bigvee_{\substack{t=1, \dots, T: \\ (\text{an}_{\text{mw}}^{\overline{O}_t}(i) \cap \text{cAn}^o(O_t)) \neq \emptyset}} \frac{b_{j_{t,i}^{\text{mw}} i}}{b_{j_{t,i}^{\text{mw}} i_t}} X_{i_t} \vee \bigvee_{j \in \text{An}_{\text{nmw}}^{\overline{O}}(i)} b_{ji} Z_j. \end{aligned} \quad (5.12)$$

*Proof.* Assume wlog that  $\mathbb{P}(\Omega^o) > 0$ . The following holds  $\mathbb{P}$ -a.s. on  $\Omega^o$ . For the existence of the nodes  $i_t$  see the proof of Proposition 5.13. Observe that, as the first equality in (4.4) holds, we can replace in (5.11)  $\text{an}_{\text{low}}^U(i)$  by  $\text{an}(i) \cap \overline{O}$ ; i.e., we have

$$X_i = \bigvee_{t=1, \dots, T} \bigvee_{k \in \text{an}(i) \cap \overline{O}_t} \frac{b_{ki}}{b_{kk}} X_k \vee \bigvee_{j \in \text{An}_{\text{nmw}}^{\overline{O}}(i)} b_{ji} Z_j.$$

Now let  $t \in \{1, \dots, T\}$  such that  $\text{an}_{\text{mw}}^{\overline{O}_t}(i) \cap \text{cAn}^o(O_t) = \emptyset$ , and let  $k \in \text{an}(i) \cap \overline{O}_t$ . Observe from (5.7) that  $\text{cAn}^o(O_t) \subseteq \text{An}^o(k) \subseteq \text{An}(k)$ ; hence,  $\text{cAn}^o(O_t) \subseteq \text{An}(i)$ . Furthermore, note for all  $j \in \text{cAn}^o(O_t)$  that  $j \in \text{an}_{\text{nmw}}^{\{k\}}(i)$ , since  $\text{an}_{\text{mw}}^{\overline{O}_t}(i) \cap \text{cAn}^o(O_t) = \emptyset$ . Altogether, also using (5.9) and (2.7), we obtain

$$\frac{b_{ki}}{b_{kk}} X_k = \frac{b_{ki}}{b_{kk}} \left( \bigvee_{j \in \text{cAn}^o(O_t)} b_{jk} Z_j \right) < \bigvee_{j \in \text{cAn}^o(O_t)} b_{ji} Z_j \leq \bigvee_{j \in \text{An}(i)} b_{ji} Z_j = X_i.$$

Thus we can omit those terms in (5.11) which correspond to a weakly connected component  $C_t$  satisfying  $\text{an}_{\text{mw}}^{C_t \cap \overline{O}}(i) \cap \text{cAn}^o(O_t) = \emptyset$  and obtain with (5.10)

$$\begin{aligned} X_i &= \bigvee_{\substack{t=1, \dots, T: \\ (\text{an}_{\text{mw}}^{C_t \cap \overline{O}}(i) \cap \text{cAn}^o(O_t)) \neq \emptyset}} \left( \bigvee_{k \in \text{an}(i) \cap \overline{O}_t} \frac{b_{j_{t,i}^{\text{mw}} k} b_{ki}}{b_{j_{t,i}^{\text{mw}} j_{t,i}^{\text{mw}}} b_{kk}} \right) X_{j_{t,i}^{\text{mw}}}^u \vee \bigvee_{j \in \text{An}_{\text{nmw}}^{\overline{O}}(i)} b_{ji} Z_j \\ &= \bigvee_{\substack{t=1, \dots, T: \\ (\text{an}_{\text{mw}}^{C_t \cap \overline{O}}(i) \cap \text{cAn}^o(O_t)) \neq \emptyset}} \left( \bigvee_{k \in \text{an}(i) \cap \overline{O}_t} \frac{b_{j_{t,i}^{\text{mw}} k} b_{ki}}{b_{j_{t,i}^{\text{mw}} i_t} b_{kk}} \right) X_{i_t} \vee \bigvee_{j \in \text{An}_{\text{nmw}}^{\overline{O}}(i)} b_{ji} Z_j. \end{aligned}$$

Since  $j_t^{\text{mw}} \in \text{an}_{\text{mw}}^{\overline{O}_t}(i)$ , applying (2.6) finishes the proof.  $\square$

## 6. Regular conditional distributions of a max-linear graphical model

Throughout this section  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  is a ML graphical model with ML coefficient matrix  $B = (b_{ij})_{d \times d}$ . Also denote again by  $(\Omega, \mathcal{A}, \mathbb{P})$  the probability space on which  $\mathbf{X}$  is defined. For  $A, O \subseteq V$  our goal is to provide an explicit formula for a regular conditional distribution function  $F_{\mathbf{X}_A | \mathbf{X}_O}$  of  $\mathbf{X}_A$  given  $\mathbf{X}_O$ .

Recall the following facts about regular conditional distributions (see for example Chapter 8.3 of [4]).

**Definition 6.1.** For  $d \in \mathbb{N}$  we denote by  $\mathcal{B}(\mathbb{R}_+^d)$  the Borel  $\sigma$ -algebra on  $\mathbb{R}_+^d$ . Let  $\mathbf{Y} = (Y_1, \dots, Y_m)$  and  $\mathbf{Z} = (Z_1, \dots, Z_n)$  be random vectors on  $(\Omega, \mathcal{A}, \mathbb{P})$  with values in  $(\mathbb{R}_+^m, \mathcal{B}(\mathbb{R}_+^m))$  and  $(\mathbb{R}_+^n, \mathcal{B}(\mathbb{R}_+^n))$ , respectively. We denote by  $\mathbb{P}_{\mathbf{Z}}$  the image measure of  $\mathbb{P}$  under  $\mathbf{Z}$ . A map  $\kappa : \mathbb{R}_+^n \times \mathcal{B}(\mathbb{R}_+^m) \rightarrow [0, 1]$  is called *regular conditional distribution of  $\mathbf{Y}$  given  $\mathbf{Z}$*  if the following conditions are satisfied:

- (a) the function  $\mathbf{z} \mapsto \kappa(\mathbf{z}, B)$  is measurable for every fixed  $B \in \mathcal{B}(\mathbb{R}_+^m)$ ,
- (b)  $B \mapsto \kappa(\mathbf{z}, B)$  is a probability measure on  $(\mathbb{R}_+^m, \mathcal{B}(\mathbb{R}_+^m))$  for every fixed  $\mathbf{z} \in \mathbb{R}_+^n$ , and
- (c)  $\mathbb{P}(\mathbf{Z} \in A, \mathbf{Y} \in B) = \int_A \kappa(\mathbf{z}, B) \mathbb{P}_{\mathbf{Z}}(d\mathbf{z})$  for every  $A \in \mathcal{B}(\mathbb{R}_+^n)$  and  $B \in \mathcal{B}(\mathbb{R}_+^m)$ .

The function  $\mathbf{y} \mapsto F_{\mathbf{Y} | \mathbf{Z}}(\mathbf{y} | \mathbf{z}) = \kappa(\mathbf{z}, (\mathbf{0}, \mathbf{y}])$  on  $\mathbb{R}_+^m$  for  $\mathbf{z} \in \mathbb{R}_+^n$  is called *regular conditional distribution function of  $\mathbf{Y}$  given  $\mathbf{Z} = \mathbf{z}$* . (Note that  $F_{\mathbf{Y} | \mathbf{Z}}$  is uniquely defined up to almost sure equality with respect to  $\mathbb{P}_{\mathbf{Z}}$ .)  $\square$

Define  $B_O$  as the  $|\text{An}(O)| \times |O|$ -submatrix of  $B$  with entries  $b_{ji}$  for  $j \in \text{An}(O)$  and  $i \in O$ . Observe from Definition 6.1(c) that it suffices to specify  $F_{\mathbf{X}_A | \mathbf{X}_O}(\mathbf{y}_A | \mathbf{x}_O)$  only for  $\mathbf{x}_O \in \text{supp}(\mathbf{X}_O) = \{\mathbf{x} \in \mathbb{R}_+^{|\mathcal{O}|} : \mathbf{x} = \mathbf{z} \odot B_O \text{ for all } \mathbf{z} \in \mathbb{R}_+^{|\text{An}(O)|}\}$  (cf. (2.5)) and  $\mathbf{y}_A \in \mathbb{R}_+^{|\mathcal{A}|}$ . Also note from the definition of the upper bounds in (5.1) and Definition 5.1(a) that the images of the finitely many different order sets under  $\mathbf{X}_O$  are disjoint sets; i.e., for every  $\mathbf{x}_O \in \text{supp}(\mathbf{X}_O)$  there exists exactly one order set  $\Omega^o$  such that  $\mathbf{x}_O \in \mathbf{X}_O(\Omega^o)$ . We show that  $F_{\mathbf{X}_A | \mathbf{X}_O}(\mathbf{y}_A | \mathbf{x}_O)$  depends on which order set corresponds to  $\mathbf{x}_O$ .

For a general max-linear  $\mathbf{X} = (X_1, \dots, X_d)$  as in (2.4) a regular conditional distribution of  $(Z_1, \dots, Z_d)$  given  $\mathbf{X}$  has been computed in [12], mainly for simulation purposes in the framework of prediction. We reduce a regular conditional distribution function  $F_{\mathbf{X}_A | \mathbf{X}_O}$  of  $\mathbf{X}_A$  given  $\mathbf{X}_O$  to the situation considered by these authors such that we can use their results. In contrast to the case of general max-linear models we can use the DAG  $\mathcal{D}$ , the order DAG  $\mathcal{D}^o$  corresponding to the value of  $\mathbf{X}_O$  on which we condition, and other embedded graph structures. This results in a reduced form for the regular conditional distribution function of  $\mathbf{X}_A$  given  $\mathbf{X}_O$ .

We start with an example.

**Example 6.2.** [Continuation of Examples 2.1, 2.8, 4.5, 4.10, 5.15: computing a regular conditional distribution]

For  $i = 1, \dots, d$  assume that the noise variable  $Z_i$  has distribution function  $F_{Z_i}$  and density  $f_{Z_i}$ . For this simple example we can compute regular conditional distributions in an informal way, and we do this for  $A = \{1, 4\}$  and  $O = \{2, 3\}$ . A precise mathematical justification will be given later on in this section.

Using the independence of the noise variables we obtain for  $(y_1, y_4) \in \mathbb{R}_+^2$  and  $(x_2, x_3) \in \text{supp}(\mathbf{X}_O) = \mathbb{R}_+^2$ ,

$$\begin{aligned} & F_{(X_1, X_4)|(X_2, X_3)}(y_1, y_4 \mid x_2, x_3) \\ &= \mathbb{P}(X_1 \leq y_1, X_4 \leq y_4 \mid X_2 = x_2, X_3 = x_3) \\ &= \mathbb{P}(c_1^1 Z_1 \leq y_1, c_2^4 x_2 \vee c_3^4 x_3 \vee c_4^4 Z_4 \leq y_4 \mid X_2 = c_1^2 Z_1 \vee c_2^2 Z_2 = x_2, X_3 = c_1^3 Z_1 \vee c_3^3 Z_3 = x_3) \\ &= \mathbf{1}_{(0, y_4]}(c_2^4 x_2 \vee c_3^4 x_3) F_{Z_4}\left(\frac{y_4}{c_4^4}\right) \mathbb{P}(X_1 \leq y_1 \mid X_2 = x_2, X_3 = x_3) \\ &= \mathbf{1}_{(0, y_4]} \left( \frac{b_{24}}{b_{22}} x_2 \vee \frac{b_{34}}{b_{33}} x_3 \right) F_{Z_4}\left(\frac{y_4}{b_{44}}\right) F_{X_1|(X_2, X_3)}(y_1 \mid x_2, x_3), \end{aligned}$$

where  $F_{X_1|(X_2, X_3)}$  denotes a regular conditional distribution function of  $X_1$  given  $(X_2, X_3)$ . In the following we provide an explicit formula for  $F_{X_1|(X_2, X_3)}$ .

Observe that we have three different order sets, namely  $\Omega_1^o = \{\omega \in \Omega : b_{13}X_2(\omega) = b_{12}X_3(\omega)\}$ ,  $\Omega_2^o = \{\omega \in \Omega : b_{13}X_2(\omega) < b_{12}X_3(\omega)\}$ , and  $\Omega_3^o = \{\omega \in \Omega : b_{13}X_2(\omega) > b_{12}X_3(\omega)\}$  with order DAGs  $\mathcal{D}_1^o = (\{1, 2, 3\}, \{(1, 2), (1, 3)\})$ ,  $\mathcal{D}_2^o = (\{1, 2, 3\}, \{(1, 2)\})$ , and  $\mathcal{D}_3^o = (\{1, 2, 3\}, \{(1, 3)\})$ .

Firstly, let  $(x_2, x_3) \in (\mathbf{X}_O(\Omega_1^o) \cap \text{supp}(\mathbf{X}_O)) = \{(x_2, x_3) \in \mathbb{R}_+^2 : b_{13}x_2 = b_{12}x_3\}$ . Then from (5.10) we obtain  $\mathbb{P}$ -a.s. on  $\Omega_1^o$ ,

$$X_1 = \frac{b_{11}}{b_{12}} X_2 = \frac{b_{11}}{b_{13}} X_3.$$

Hence, we have

$$F_{X_1|(X_2, X_3)}(y_1 \mid x_2, x_3) = \mathbb{P}\left(X_1 \leq y_1 \mid X_2 = \frac{b_{12}}{b_{13}} X_3 = x_2\right) = \mathbb{P}\left(X_1 \leq y_1 \mid X_1 = \frac{b_{11}}{b_{12}} x_2\right) = \mathbf{1}_{(0, y_1]} \left(\frac{b_{11}}{b_{12}} x_2\right).$$

Secondly, let  $(x_2, x_3) \in (\mathbf{X}_O(\Omega_2^o) \cap \text{supp}(\mathbf{X}_O)) = \{(x_2, x_3) \in \mathbb{R}_+^2 : b_{13}x_2 < b_{12}x_3\}$ . Using (5.5), the independence of the noise variables, and the fact that by Theorem 2.11(a) the node  $j(2)$  as in (2.8) is  $\mathbb{P}$ -a.s. unique, we obtain

$$\begin{aligned} & F_{X_1|(X_2, X_3)}(y_1, y_4 \mid x_2, x_3) = \mathbb{P}(b_{11}Z_1 \leq y_1 \mid b_{12}Z_1 \vee b_{22}Z_2 = x_2, b_{33}Z_3 = x_3) \\ &= \mathbb{P}(b_{11}Z_1 \leq y_1 \mid b_{12}Z_1 \vee b_{22}Z_2 = x_2) = \lim_{\varepsilon \downarrow 0} \frac{\mathbb{P}(b_{11}Z_1 \leq y_1, (b_{12}Z_1 \vee b_{22}Z_2) \in [x_2, x_2 + \varepsilon))}{\mathbb{P}(X_2 \in [x_2, x_2 + \varepsilon))} \\ &= \lim_{\varepsilon \downarrow 0} \frac{\mathbb{P}(b_{11}Z_1 \leq y_1, \{b_{12}Z_1 \in [x_2, x_2 + \varepsilon), b_{22}Z_2 \leq x_2 \text{ or } b_{12}Z_1 \leq x_2, b_{22}Z_2 \in [x_2, x_2 + \varepsilon)\})}{\mathbb{P}(X_2 \in [x_2, x_2 + \varepsilon))} \\ &= \lim_{\varepsilon \downarrow 0} \frac{\frac{1}{\varepsilon} \mathbb{P}(b_{11}Z_1 \leq y_1, b_{12}Z_1 \in [x_2, x_2 + \varepsilon), b_{22}Z_2 \leq x_2) + \frac{1}{\varepsilon} \mathbb{P}(b_{11}Z_1 \leq y_1, b_{12}Z_1 \leq x_2, b_{22}Z_2 \in [x_2, x_2 + \varepsilon))}{\frac{1}{\varepsilon} \mathbb{P}(X_2 \in [x_2, x_2 + \varepsilon))} \\ &= \frac{\frac{1}{b_{12}} \mathbf{1}_{(0, \frac{y_1}{b_{11}}]} \left(\frac{x_2}{b_{12}}\right) f_{Z_1}\left(\frac{x_2}{b_{12}}\right) F_{Z_2}\left(\frac{x_2}{b_{22}}\right) + \frac{1}{b_{22}} f_{Z_2}\left(\frac{x_2}{b_{22}}\right) F_{Z_1}\left(\frac{y_1}{b_{11}} \wedge \frac{x_2}{b_{12}}\right)}{f_{X_2}(x_2)} \\ &= \frac{\frac{1}{b_{12}} \mathbf{1}_{(0, \frac{y_1}{b_{11}}]} \left(\frac{x_2}{b_{12}}\right) f_{Z_1}\left(\frac{x_2}{b_{12}}\right) F_{Z_2}\left(\frac{x_2}{b_{22}}\right) + \frac{1}{b_{22}} f_{Z_2}\left(\frac{x_2}{b_{22}}\right) F_{Z_1}\left(\frac{x_2}{b_{12}} \wedge \frac{y_1}{b_{11}}\right)}{\frac{1}{b_{12}} f_{Z_1}\left(\frac{x_2}{b_{12}}\right) F_{Z_2}\left(\frac{x_2}{b_{22}}\right) + \frac{1}{b_{22}} f_{Z_2}\left(\frac{x_2}{b_{22}}\right) F_{Z_1}\left(\frac{x_2}{b_{12}}\right)}. \end{aligned}$$

By the symmetry of the graph a regular conditional distribution function of  $X_1$  given  $(X_2, X_3) = (x_2, x_3)$  corresponding to the third case,  $(x_2, x_3) \in (\mathbf{X}_O(\Omega_3^o) \cap \text{supp}(\mathbf{X}_O)) = \{(x_2, x_3) \in \mathbb{R}_+^2 : b_{13}x_2 > b_{12}x_3\}$ , can be obtained by reversing the roles of nodes 2 and 3 in the second case.  $\square$

To prove that a function  $F_{\mathbf{X}_A|\mathbf{X}_O} : \mathbb{R}_+^{[A]} \times \text{supp}(\mathbf{X}_O) \rightarrow [0, 1]$  is a regular conditional distribution function of  $\mathbf{X}_A$  given  $\mathbf{X}_O$ , we have to verify the properties of Definition 6.1. In particular, to verify (c) it suffices to show that

$$\mathbb{P}(\mathbf{X}_A \leq \mathbf{y}_A, \mathbf{X}_O \leq \mathbf{w}_O) = \int_{(\mathbf{0}, \mathbf{w}_O]} F_{\mathbf{X}_A|\mathbf{X}_O}(\mathbf{y}_A | \mathbf{x}_O) \mathbb{P}_{\mathbf{X}_O}(d\mathbf{x}_O) \quad (6.1)$$

for all  $\mathbf{y}_A \in \mathbb{R}_+^{[A]}$  and  $\mathbf{w}_O \in \mathbb{R}_+^{[O]}$ . We first present our concept of proof.

Assume that there exist  $n \in \mathbb{N}$  different order sets, say  $\Omega_1^o, \dots, \Omega_n^o$ . Since the image sets  $\mathbf{X}_O(\Omega_s^o)$  for  $s = 1, \dots, n$  are disjoint, we can write for  $\mathbf{y}_A \in \mathbb{R}_+^{[A]}$  and  $\mathbf{x}_O \in \text{supp}(\mathbf{X}_O)$ ,

$$F_{\mathbf{X}_A|\mathbf{X}_O}(\mathbf{y}_A | \mathbf{x}_O) = \sum_{s=1}^n F_{\mathbf{X}_A|\mathbf{X}_O}(\mathbf{y}_A | \mathbf{x}_O) \mathbf{1}_{\mathbf{X}_O(\Omega_s^o)}(\mathbf{x}_O).$$

From this we can conclude

$$\int_{(\mathbf{0}, \mathbf{w}_O]} F_{\mathbf{X}_A|\mathbf{X}_O}(\mathbf{y}_A | \mathbf{x}_O) \mathbb{P}_{\mathbf{X}_O}(d\mathbf{x}_O) = \sum_{s=1}^n \int_{(\mathbf{0}, \mathbf{w}_O]} F_{\mathbf{X}_A|\mathbf{X}_O}(\mathbf{y}_A | \mathbf{x}_O) \mathbf{1}_{\mathbf{X}_O(\Omega_s^o)}(\mathbf{x}_O) \mathbb{P}_{\mathbf{X}_O}(\mathbf{x}_O).$$

Furthermore, since the order sets  $\Omega_s^o$  for  $s = 1, \dots, n$  are disjoint, we also have for  $\mathbf{w}_O \in \mathbb{R}_+^{[O]}$ ,

$$\mathbb{P}(\mathbf{X}_A \leq \mathbf{y}_A, \mathbf{X}_O \leq \mathbf{w}_O) = \sum_{s=1}^n \mathbb{P}(\mathbf{X}_A \leq \mathbf{y}_A, \mathbf{X}_O \in ((\mathbf{0}, \mathbf{w}_O] \cap \mathbf{X}_O(\Omega_s^o))).$$

Hence, (6.1) holds, if we can show for  $s = 1, \dots, n$  that

$$\mathbb{P}(\mathbf{X}_A \leq \mathbf{y}_A, \mathbf{X}_O \in ((\mathbf{0}, \mathbf{w}_O] \cap \mathbf{X}_O(\Omega_s^o))) = \int_{(\mathbf{0}, \mathbf{w}_O]} F_{\mathbf{X}_A|\mathbf{X}_O}(\mathbf{y}_A | \mathbf{x}_O) \mathbf{1}_{\mathbf{X}_O(\Omega_s^o)}(\mathbf{x}_O) \mathbb{P}_{\mathbf{X}_O}(\mathbf{x}_O)$$

for all  $\mathbf{y}_A \in \mathbb{R}_+^{[A]}$  and  $\mathbf{w}_O \in \mathbb{R}_+^{[O]}$ .

These considerations justify our proceeding in the following. For some fixed but arbitrary order set  $\Omega^o$  we present a function  $F_{\mathbf{X}_A|\mathbf{X}_O} : \mathbb{R}_+^{[A]} \times (\mathbf{X}_O(\Omega^o) \cap \text{supp}(\mathbf{X}_O)) \rightarrow [0, 1]$  and show that the properties (a) and (b) from Definition 6.1 as well as

$$\mathbb{P}(\mathbf{X}_A \leq \mathbf{y}_A, \mathbf{X}_O \in ((\mathbf{0}, \mathbf{w}_O] \cap \mathbf{X}_O(\Omega^o))) = \int_{(\mathbf{0}, \mathbf{w}_O]} F_{\mathbf{X}_A|\mathbf{X}_O}(\mathbf{y}_A | \mathbf{x}_O) \mathbf{1}_{\mathbf{X}_O(\Omega^o)}(\mathbf{x}_O) \mathbb{P}_{\mathbf{X}_O}(\mathbf{x}_O) \quad (6.2)$$

for all  $\mathbf{y}_A \in \mathbb{R}_+^{[A]}$  and  $\mathbf{w}_O \in \mathbb{R}_+^{[O]}$  hold.

For this  $\Omega^o$  let  $\overline{O}$  be the corresponding extended  $O$ -set as defined in (5.7). For  $i \in V$  recall from Definition 2.9(a) the set  $\text{An}_{\text{nmw}}^{\overline{O}}(i)$  of ancestors of  $i$  which are not max-weighted by  $\overline{O}$ , and define

$$S := \text{An}_{\text{nmw}}^{\overline{O}}(A \setminus \overline{O}) = \bigcup_{i \in A \setminus \overline{O}} \text{An}_{\text{nmw}}^{\overline{O}}(i). \quad (6.3)$$

Proposition 5.17 allows us now to reduce a regular conditional distribution function  $F_{\mathbf{X}_A|\mathbf{X}_O}$  to a regular conditional distribution function of  $\mathbf{Z}_S$  given  $\mathbf{X}_O$ , which we will calculate subsequently. The following is the main structural result of this section. Here and in what follows, for arbitrary  $a_i \geq 0$  we set  $\prod_{i \in \emptyset} a_i = 1$ .

**Theorem 6.3.** *Let  $\Omega^o$  be some order set with order DAG  $\mathcal{D}^o$ . Assume that the weakly connected components of  $\mathcal{D}^o$  have node sets  $C_1, \dots, C_T$ , and recall (5.4) as well as (5.7). For  $i \in \overline{O}^c$  and  $t = 1, \dots, T$  let  $j_{t,i}^{\text{mw}} \in \text{an}_{\text{mw}}^{\overline{O}_t}(i) \cap \text{cAn}^o(O_t)$  and  $i_t \in O_t$  be arbitrary but fixed nodes as in Proposition 5.17. For  $\mathbf{x}_O \in \mathbf{X}_O(\Omega^o)$  we denote the corresponding realized upper bounds in (5.1) by  $x_i^u$ . If the set  $S$  as in (6.3) is non-empty, let  $F_{\mathbf{Z}_S|\mathbf{X}_O}$  be a regular conditional distribution function of  $\mathbf{Z}_S$  given  $\mathbf{X}_O$ . Then for  $\mathbf{y}_A \in \mathbb{R}_+^{[A]}$  and  $\mathbf{x}_O \in (\mathbf{X}_O(\Omega^o) \cap \text{supp}(\mathbf{X}_O))$ ,*

$$\begin{aligned} & F_{\mathbf{X}_A|\mathbf{X}_O}(\mathbf{y}_A | \mathbf{x}_O) \\ &= \prod_{i \in A \setminus \overline{O}} \mathbf{1}_{[0, y_i]}(x_i^u) \prod_{i \in A \setminus \overline{O}} \mathbf{1}_{[0, y_i]} \left( \bigvee_{\substack{t=1, \dots, T: \\ (\text{an}_{\text{mw}}^{\overline{O}_t}(i) \cap \text{cAn}^o(O_t)) \neq \emptyset}} \frac{b_{j_{t,i}^{\text{mw}}}}{b_{j_{t,i}^{\text{mw}} j_{t,i}^{\text{mw}}}^{\text{mw}}} x_{j_{t,i}^{\text{mw}}}^u \right) F_{\mathbf{Z}_S|\mathbf{X}_O}(\mathbf{z}_S | \mathbf{x}_O) \\ &= \prod_{t=1, \dots, T} \prod_{i \in A \setminus \overline{O}_t} \mathbf{1}_{[0, y_i]} \left( \frac{b_{j_{t,i}^{\text{mw}}}}{b_{j_{t,i}^{\text{mw}} i_t}^{\text{mw}}} x_{i_t} \right) \prod_{i \in A \setminus \overline{O}} \mathbf{1}_{[0, y_i]} \left( \bigvee_{\substack{t=1, \dots, T: \\ (\text{an}_{\text{mw}}^{\overline{O}_t}(i) \cap \text{cAn}^o(O_t)) \neq \emptyset}} \frac{b_{j_{t,i}^{\text{mw}}}}{b_{j_{t,i}^{\text{mw}} i_t}^{\text{mw}}} x_{i_t} \right) F_{\mathbf{Z}_S|\mathbf{X}_O}(\mathbf{z}_S | \mathbf{x}_O), \end{aligned}$$

where  $z_j = \bigwedge_{i \in \text{De}_{\text{nmw}}^{\overline{O}}(j) \cap A \setminus \overline{O}} \frac{y_i}{b_{ji}}$  for  $j \in S$ . In case that  $S$  is empty, a regular conditional distribution function of  $\mathbf{X}_A$  given  $\mathbf{X}_O$  is given by the formulas obtained from the above by replacing  $F_{\mathbf{Z}_S|\mathbf{X}_O}$  by 1.

*Proof.* Assume wlog that  $\mathbb{P}(\Omega^o) > 0$ . First, observe from (5.10) that the two formulas are equivalent. By the definition of  $F_{\mathbf{X}_A|\mathbf{X}_O}(\mathbf{y}_A | \mathbf{x}_O)$ , the properties (a) and (b) from Definition 6.1 are satisfied. Thus it remains to show that (6.2) holds. For  $\mathbf{w}_O \in \mathbb{R}_+^{[O]}$  and  $\mathbf{y}_A \in \mathbb{R}_+^{[A]}$  define

$$\mathbb{P}^o(\mathbf{X}_A \leq \mathbf{y}_A) := \mathbb{P}(\mathbf{X}_A \leq \mathbf{y}_A, \mathbf{X}_O \in ((\mathbf{0}, \mathbf{w}_O] \cap \mathbf{X}_O(\Omega^o))).$$

Using (5.10) and (5.12) we obtain

$$\begin{aligned} \mathbb{P}^o(\mathbf{X}_A \leq \mathbf{y}_A) &= \mathbb{P}(X_i \leq y_i \forall i \in A \cap \overline{O}, X_i \leq y_i \forall i \in A \setminus \overline{O}) \\ &= \mathbb{P}^o\left(\frac{b_{j_t i}}{b_{j_t i_t}} X_{i_t} \leq y_i \forall t \in \{1, \dots, T\} : A \cap \overline{O}_t \neq \emptyset \text{ and } i \in A \cap \overline{O}_t, \right. \\ &\quad \bigvee_{\substack{t=1, \dots, T: \\ (\text{an}_{\text{mw}}^{\overline{O}_t}(i) \cap \text{cAn}^o(O_t)) \neq \emptyset}} \frac{b_{j_t i}}{b_{j_t i_t}} X_{i_t} \vee \bigvee_{j \in \text{An}_{\text{nmw}}^{\overline{O}}(i)} b_{ji} Z_j \leq y_i \forall i \in A \setminus \overline{O}). \end{aligned}$$

For notational convenience we define the sets

$$E_1 := \{t \in \{1, \dots, T\} : A \cap \overline{O}_t \neq \emptyset\} \quad \text{and} \quad E_2 := \{t \in \{1, \dots, T\} : \text{an}_{\text{mw}}^{\overline{O}_t}(A \setminus \overline{O}) \cap \text{cAn}^o(O_t) \neq \emptyset\}.$$

We simplify both components in the above probability, where the first is obvious. For the second we use Lemma A.3 with  $E = A \setminus \overline{O}$ . Hence,

$$\mathbb{P}^o(\mathbf{X}_A \leq \mathbf{y}_A) = \mathbb{P}^o(X_{i_t} \leq \bigwedge_{i \in A \cap \overline{O}_t} \frac{b_{j_t i}}{b_{j_t i_t}} y_i \forall t \in E_1, X_{i_t} \leq \bigwedge_{i \in \text{de}_{\text{mw}}^{\overline{O}_t}(\text{cAn}^o(O_t)) \cap A \setminus \overline{O}} \frac{b_{j_t i}}{b_{j_t i_t}} y_i \forall t \in E_2, \mathbf{Z}_S \leq \mathbf{z}_S).$$

Similar reformulations as above for  $x_{i_t}$  instead of  $X_{i_t}$  yield

$$\begin{aligned} F_{\mathbf{X}_A|\mathbf{X}_O}(\mathbf{y}_A | \mathbf{x}_O) &= \prod_{t \in E_1} \mathbf{1}_{[0, \bigwedge_{i \in A \cap \overline{O}_t} \frac{b_{j_t i}}{b_{j_t i_t}} y_i]}(x_{i_t}) \prod_{t \in E_2} \mathbf{1}_{[0, \bigwedge_{i \in \text{de}_{\text{mw}}^{\overline{O}_t}(\text{cAn}^o(O_t)) \cap A \setminus \overline{O}} \frac{b_{j_t i}}{b_{j_t i_t}} y_i]}(x_{i_t}) F_{\mathbf{Z}_S|\mathbf{X}_O}(\mathbf{z}_S | \mathbf{x}_O). \end{aligned}$$

Integration shows that the property of Definition 6.1(c) holds, since it holds for  $F_{\mathbf{Z}_S|\mathbf{X}_O}$ , which we have assumed to be a regular conditional distribution function. Hence, (6.2) holds.  $\square$

Observe that in the proof of Theorem 6.3 for nodes of  $A \setminus \overline{O}$  we have used representation (5.12). Similarly, we may also apply the representations from Corollary 5.14. The following remark shows the regular conditional distribution function of  $\mathbf{X}_A$  given  $\mathbf{X}_O$  corresponding to these representations. Of course all formulas are equivalent.

**Remark 6.4.** Assume the same conditions as in Theorem 6.3 hold. For  $t = 1, \dots, T$  let  $i_t \in O_t$  and  $j_t \in \text{cAn}^o(O_t)$  be arbitrary but fixed nodes. Then, if the set  $S$  as in (6.3) is non-empty, for  $\mathbf{y}_A \in \mathbb{R}_+^{[A]}$  and  $\mathbf{x}_O \in (\mathbf{X}_O(\Omega^o) \cap \text{supp}(\mathbf{X}_O))$ ,

$$\begin{aligned} F_{\mathbf{X}_A|\mathbf{X}_O}(\mathbf{y}_A | \mathbf{x}_O) &= \prod_{i \in A \cap \overline{O}} \mathbf{1}_{[0, y_i]}(x_i^u) \prod_{i \in A \setminus \overline{O}} \mathbf{1}_{[0, y_i]} \left( \bigvee_{k \in \text{an}_{\text{low}}^{\overline{O}}(i)} \frac{b_{ki}}{b_{kk}} x_k^u \right) F_{\mathbf{Z}_S|\mathbf{X}_O}(\mathbf{z}_S | \mathbf{x}_O) \\ &= \prod_{t=1, \dots, T} \prod_{i \in A \cap \overline{O}_t} \mathbf{1}_{[0, y_i]} \left( \frac{b_{j_t i}}{b_{j_t i_t}} x_{i_t} \right) \prod_{i \in A \setminus \overline{O}} \mathbf{1}_{[0, y_i]} \left( \bigvee_{t=1, \dots, T} \left( \bigvee_{k \in \text{an}_{\text{low}}^{\overline{O}_t}(i) \cap C_t} \frac{b_{j_t k} b_{ki}}{b_{j_t j_t} b_{kk}} \right) x_{j_t}^u \right) F_{\mathbf{Z}_S|\mathbf{X}_O}(\mathbf{z}_S | \mathbf{x}_O) \\ &= \prod_{t=1, \dots, T} \prod_{i \in A \cap \overline{O}_t} \mathbf{1}_{[0, y_i]} \left( \frac{b_{j_t i}}{b_{j_t i_t}} x_{i_t} \right) \prod_{i \in A \setminus \overline{O}} \mathbf{1}_{[0, y_i]} \left( \bigvee_{t=1, \dots, T} \left( \bigvee_{k \in \text{an}_{\text{low}}^{\overline{O}_t}(i) \cap C_t} \frac{b_{j_t k} b_{ki}}{b_{j_t i_t} b_{kk}} \right) x_{i_t} \right) F_{\mathbf{Z}_S|\mathbf{X}_O}(\mathbf{z}_S | \mathbf{x}_O). \end{aligned}$$

In case that  $S$  is empty, a regular conditional distribution function of  $\mathbf{X}_A$  given  $\mathbf{X}_O$  is given by the formulas obtained from the above by replacing  $F_{\mathbf{Z}_S|\mathbf{X}_O}$  by 1.  $\square$



In the next step our intention is to find a regular conditional distribution function of  $\mathbf{Z}_S$  given  $\mathbf{X}_O = \mathbf{x}_O$  for  $\mathbf{x}_O \in (\mathbf{X}_O(\Omega^o) \cap \text{supp}(\mathbf{X}_O))$ . First, for  $\mathbf{x}_O \in \mathbb{R}_+^{|\mathcal{O}|}$  we reduce a regular conditional distribution function of  $\mathbf{Z}_S$  given  $\mathbf{X}_O = \mathbf{x}_O$  to a regular conditional distribution function of a subvector of  $\mathbf{Z}_S$  given  $\mathbf{X}_O = \mathbf{x}_O$ . For this purpose recall from Theorem 5.9 that  $\mathbf{X}_O$  is a function of  $\mathbf{Z}_{\text{An}(O)}$ . Thus the independence of the noise variables suggests a partition of the set  $S = \text{An}_{\text{nmw}}^{\overline{\mathcal{O}}}(A \setminus \overline{\mathcal{O}})$  into the sets

$$S_1 := \text{An}_{\text{nmw}}^{\overline{\mathcal{O}}}(A \setminus \overline{\mathcal{O}}) \setminus \text{An}(O) = \text{An}_{\text{nmw}}^{\overline{\mathcal{O}}}(A \setminus \overline{\mathcal{O}}) \setminus \text{an}(O), \quad (6.4)$$

$$S_2 := \text{An}_{\text{nmw}}^{\overline{\mathcal{O}}}(A \setminus \overline{\mathcal{O}}) \cap \text{An}(O) = \text{An}_{\text{nmw}}^{\overline{\mathcal{O}}}(A \setminus \overline{\mathcal{O}}) \cap \text{an}(O), \quad (6.5)$$

where we have the second equality in each definition, since every path from a node in  $O$  to  $i \in A \setminus \overline{\mathcal{O}}$  is max-weighted by  $\overline{\mathcal{O}}$ . This partition determines a factorization of a regular conditional distribution function  $F_{\mathbf{Z}_S|\mathbf{x}_O}$  of  $\mathbf{Z}_S$  given  $\mathbf{X}_O$ . As we present the following lemma in the context of Theorem 6.3, we work under the very same conditions. Indeed, the result would also hold under weaker conditions.

**Lemma 6.5.** *Assume the same conditions as in Theorem 6.3 hold. Suppose also that the set  $S$  as in (6.3) is non-empty. Let  $S_1$  and  $S_2$  be as in (6.4) and (6.5). If  $S_2$  is non-empty, let  $F_{\mathbf{Z}_{S_2}|\mathbf{x}_O}$  be a regular conditional distribution function of  $\mathbf{Z}_{S_2}$  given  $\mathbf{X}_O$ . Then for  $\mathbf{z}_S \in \mathbb{R}_+^{|\mathcal{S}|}$  and  $\mathbf{x}_O \in \mathbb{R}_+^{|\mathcal{O}|}$ ,*

$$F_{\mathbf{Z}_S|\mathbf{x}_O}(\mathbf{z}_S | \mathbf{x}_O) = F_{\mathbf{Z}_{S_1}}(\mathbf{z}_{S_1})F_{\mathbf{Z}_{S_2}|\mathbf{x}_O}(\mathbf{z}_{S_2} | \mathbf{x}_O) = \prod_{j \in S_1} F_{Z_j}(z_j)F_{\mathbf{Z}_{S_2}|\mathbf{x}_O}(\mathbf{z}_{S_2} | \mathbf{x}_O).$$

*In case that  $S_2$  is empty, a regular conditional distribution function of  $\mathbf{Z}_S$  given  $\mathbf{X}_O$  is given by the formulas obtained from the above by replacing  $F_{\mathbf{Z}_{S_2}|\mathbf{x}_O}$  by 1.*

*Proof.* We have to verify the properties of Definition 6.1. By the definition of  $F_{\mathbf{Z}_S|\mathbf{x}_O}(\mathbf{z}_S | \mathbf{x}_O)$ , the properties (a) and (b) from Definition 6.1 are satisfied. Using the independence of the noise variables and the property of Definition 6.1(c) for  $F_{\mathbf{Z}_{S_2}|\mathbf{x}_O}$ , we obtain for  $\mathbf{w}_O \in \mathbb{R}_+^{|\mathcal{O}|}$  and  $\mathbf{z}_S = (\mathbf{z}_{S_1}, \mathbf{z}_{S_2}) \in \mathbb{R}_+^{|\mathcal{S}|}$ ,

$$\begin{aligned} & \int_{(\mathbf{0}, \mathbf{w}_O]} F_{\mathbf{Z}_{S_1}}(\mathbf{z}_{S_1})F_{\mathbf{Z}_{S_2}|\mathbf{x}_O}(\mathbf{z}_{S_2} | \mathbf{x}_O)\mathbb{P}_{\mathbf{x}_O}(\mathrm{d}\mathbf{x}_O) \\ &= \prod_{j \in S_1} F_{Z_j}(z_j) \int_{(\mathbf{0}, \mathbf{w}_O]} F_{\mathbf{Z}_{S_2}|\mathbf{x}_O}(\mathbf{z}_{S_2} | \mathbf{x}_O)\mathbb{P}_{\mathbf{x}_O}(\mathrm{d}\mathbf{x}_O) \\ &= \prod_{j \in S_1} F_{Z_j}(z_j)\mathbb{P}(\mathbf{X}_O \leq \mathbf{w}_O, \mathbf{Z}_{S_2} \leq \mathbf{z}_{S_2}). \end{aligned}$$

Since  $\mathbf{X}_O$  is a function of  $\mathbf{Z}_{\text{An}(O)}$  and the sets  $S_1$  and  $\text{An}(O) \cup S_2 = S_2$  are disjoint, we have by the independence of the noise variables that

$$\prod_{j \in S_1} F_{Z_j}(z_j)\mathbb{P}(\mathbf{X}_O \leq \mathbf{w}_O, \mathbf{Z}_{S_2} \leq \mathbf{z}_{S_2}) = \mathbb{P}(\mathbf{X}_O \leq \mathbf{w}_O, \mathbf{Z}_{S_1 \cup S_2} \leq \mathbf{z}_{S_1 \cup S_2}) = \mathbb{P}(\mathbf{X}_O \leq \mathbf{w}_O, \mathbf{Z}_S \leq \mathbf{z}_S).$$

Hence, property (c) is also satisfied.  $\square$

The following lemma provides a formula for a regular conditional distribution function  $F_{\mathbf{Z}_{S_2}|\mathbf{x}_O}$  of  $\mathbf{Z}_{S_2}$  given  $\mathbf{X}_O = \mathbf{x}_O$  for  $\mathbf{x}_O \in (\mathbf{X}_O(\Omega^o) \cap \text{supp}(\mathbf{X}_O))$ . Theorem 2 of [12] gives one of  $\mathbf{Z}_{\text{An}(O)}$  given  $\mathbf{X}_O$ . By setting in this result  $E = (\mathbf{0}, \mathbf{z}_{\text{An}(O)})$  for  $\mathbf{z}_{\text{An}(O)} \in \mathbb{R}_+^{|\text{An}(O)|}$  we have a regular conditional distribution function  $F_{\mathbf{Z}_{\text{An}(O)}|\mathbf{x}_O}$  of  $\mathbf{Z}_{\text{An}(O)}$  given  $\mathbf{X}_O$ . Furthermore, since  $S_2 \subseteq \text{An}(O)$ , we obtain from this theorem also a regular distribution function of  $\mathbf{Z}_{S_2}$  given  $\mathbf{X}_O$  by setting  $z_j = \infty$  for all  $j \in \text{An}(O) \setminus S_2$ . To verify that our formula for  $F_{\mathbf{Z}_{S_2}|\mathbf{x}_O}$  is indeed a version of Theorem 2 of [12] also note that their upper bounds  $\hat{z}_j$  correspond to  $x_j^u/b_{jj}$  and their  $J^{(s)}$  and  $\overline{J}^{(s)}$  for  $s = 1, \dots, r$  to  $\text{cAn}^o(O_t)$  and  $C_t$  for  $t = 1, \dots, T$  in our result, respectively.

The following lemma holds not only for  $S_2$  as in (6.5) but for some arbitrary set  $S_2 \subseteq \text{An}(O)$ . The second equality is a consequence of (5.10). In what follows, for arbitrary  $a_i \geq 0$  we set  $\sum_{i \in \emptyset} a_i = 0$ .

**Lemma 6.6.** Assume the same conditions as in Lemma 6.5 hold. Suppose also that the set  $S_2$  as in (6.5) is non-empty. For  $j \in \text{An}(O)$  define  $z_j^u := \frac{x_j^u}{b_{jj}}$ . Then for  $\mathbf{z}_{S_2} \in \mathbb{R}_+^{|S_2|}$  and  $\mathbf{x}_O \in (\mathbf{X}_O(\Omega^o) \cap \text{supp}(\mathbf{X}_O))$ ,

$$\begin{aligned}
& F_{\mathbf{z}_{S_2} | \mathbf{x}_O}(\mathbf{z}_{S_2} | \mathbf{x}_O) \\
&= \prod_{j \in S_2 \setminus (\cup_{t=1}^T \text{cAn}^o(O_t))} \frac{F_{Z_j}(z_j^u \wedge z_j)}{F_{Z_j}(z_j^u)} \prod_{\substack{t=1, \dots, T: \\ \text{cAn}^o(O_t) \cap S_2 \neq \emptyset}} \left( \sum_{j \in \text{An}^o(O_t)} z_j^u f_{Z_j}(z_j^u) \prod_{k \in \text{cAn}^o(O_t) \setminus \{j\}} F_{Z_k}(z_k^u) \right)^{-1} \\
& \quad \prod_{\substack{t=1, \dots, T: \\ \text{cAn}^o(O_t) \cap S_2 \neq \emptyset}} \left\{ \sum_{j \in \text{cAn}^o(O_t) \setminus S_2} z_j^u f_{Z_j}(z_j^u) \prod_{k \in (\text{cAn}^o(O_t) \cap S_2) \setminus \{j\}} F_{Z_k}(z_k^u \wedge z_k) \prod_{k \in \text{cAn}^o(O_t) \setminus (S_2 \cup \{j\})} F_{Z_k}(z_k^u) \right. \\
& \quad \left. + \sum_{j \in \text{cAn}^o(O_t) \cap S_2} \mathbf{1}_{(0, z_j]}(z_j^u) z_j^u f_{Z_j}(z_j^u) \prod_{k \in (\text{cAn}^o(O_t) \cap S_2) \setminus \{j\}} F_{Z_k}(z_k^u \wedge z_k) \prod_{k \in \text{cAn}^o(O_t) \setminus (S_2 \cup \{j\})} F_{Z_k}(z_k^u) \right\} \\
&= \prod_{j \in S_2 \setminus (\cup_{t=1}^T \text{cAn}^o(O_t))} \frac{F_{Z_j}(z_j^u \wedge z_j)}{F_{Z_j}(z_j^u)} \prod_{\substack{t=1, \dots, T: \\ \text{cAn}^o(O_t) \cap S_2 \neq \emptyset}} \left( \sum_{j \in \text{An}^o(O_t)} \frac{1}{b_{ji_t}} f_{Z_j}\left(\frac{x_{i_t}}{b_{ji_t}}\right) \prod_{k \in \text{cAn}^o(O_t) \setminus \{j\}} F_{Z_k}\left(\frac{x_{i_t}}{b_{ki_t}}\right) \right)^{-1} \\
& \quad \prod_{\substack{t=1, \dots, T: \\ \text{cAn}^o(O_t) \cap S_2 \neq \emptyset}} \left\{ \sum_{j \in \text{cAn}^o(O_t) \setminus S_2} \frac{1}{b_{ji_t}} f_{Z_j}\left(\frac{x_{i_t}}{b_{ji_t}}\right) \prod_{k \in (\text{cAn}^o(O_t) \cap S_2) \setminus \{j\}} F_{Z_k}\left(\frac{x_{i_t}}{b_{ki_t}} \wedge z_k\right) \prod_{k \in \text{cAn}^o(O_t) \setminus (S_2 \cup \{j\})} F_{Z_k}\left(\frac{x_{i_t}}{b_{ki_t}}\right) \right. \\
& \quad \left. + \sum_{j \in \text{cAn}^o(O_t) \cap S_2} \frac{1}{b_{ji_t}} \mathbf{1}_{(0, z_j]} \left(\frac{x_{i_t}}{b_{ji_t}}\right) f_{Z_j}\left(\frac{x_{i_t}}{b_{ji_t}}\right) \prod_{k \in (\text{cAn}^o(O_t) \cap S_2) \setminus \{j\}} F_{Z_k}\left(\frac{x_{i_t}}{b_{ki_t}} \wedge z_k\right) \prod_{k \in \text{cAn}^o(O_t) \setminus (S_2 \cup \{j\})} F_{Z_k}\left(\frac{x_{i_t}}{b_{ki_t}}\right) \right\}.
\end{aligned}$$

In the following remark we reformulate the terms in the denominator of  $F_{\mathbf{z}_{S_2} | \mathbf{x}_O}$ , giving the intuition behind. A more formal explanation can be found in Lemma 2 of [12].

**Remark 6.7.** Assume the same conditions as in Lemma 6.6 hold. Recall from (5.5) that  $\mathbb{P}$ -a.s. on  $\Omega^o$  for all  $t = 1, \dots, T$ ,

$$X_{i_t} = \bigvee_{j \in \text{cAn}^o(O_t)} b_{ji_t} Z_j.$$

Let  $t \in \{1, \dots, T\}$  be fixed. Since the noise variables are independent, the distribution function  $G_t$  of  $\bigvee_{j \in \text{cAn}^o(O_t)} b_{ji_t} Z_j$  is given by

$$G_t(y) = \mathbb{P}\left(\bigvee_{j \in \text{cAn}^o(O_t)} b_{ji_t} Z_j \leq y\right) = \prod_{j \in \text{cAn}^o(O_t)} F_{Z_j}\left(\frac{y}{b_{ji_t}}\right), \quad y \in \mathbb{R}_+.$$

By taking derivatives with respect to  $y$  we find the corresponding density  $g_t$ , namely

$$g_t(y) = \sum_{j \in \text{cAn}^o(O_t)} \frac{1}{b_{ji_t}} f_{Z_j}\left(\frac{y}{b_{ji_t}}\right) \prod_{k \in \text{cAn}^o(O_t) \setminus \{j\}} F_{Z_k}\left(\frac{y}{b_{ki_t}}\right).$$

These densities are the terms in the denominator of  $F_{\mathbf{z}_{S_2} | \mathbf{x}_O}$ . □

**Example 6.8.** [Continuation of Examples 2.1, 2.8, 4.5, 4.10, 5.15, 6.2]

We verify the regular conditional distribution functions calculated in an informal way there. From Theorem 6.3 and Lemma 6.5 we obtain for  $(y_1, y_4) \in \mathbb{R}_+^2$  and  $(x_2, x_3) \in (\mathbf{X}_O(\Omega_1^o) \cap \text{supp}(\mathbf{X}_O))$ ,

$$\begin{aligned}
F_{(X_1, X_4) | (X_2, X_3)}(y_1, y_4 | x_2, x_3) &= \mathbf{1}_{[0, y_1]} \left(\frac{b_{11}}{b_{12}} x_2\right) \mathbf{1}_{[0, y_4]} \left(\frac{b_{14}}{b_{12}} x_2\right) F_{Z_4 | (X_2, X_3)}\left(\frac{y_4}{b_{44}} | x_2, x_3\right) \\
&= \mathbf{1}_{[0, y_1]} \left(\frac{b_{11}}{b_{12}} x_2\right) \mathbf{1}_{[0, y_4]} \left(\frac{b_{14}}{b_{12}} x_2\right) F_{Z_4}\left(\frac{y_4}{b_{44}}\right).
\end{aligned}$$

Observe that by (3.5) we have for  $(x_2, x_3) \in (\mathbf{X}_O(\Omega_1^o) \cap \text{supp}(\mathbf{X}_O))$ , as  $b_{13}x_2 = b_{12}x_3$ ,

$$\frac{b_{14}}{b_{12}} x_2 = \frac{1}{b_{12}} \left( \frac{b_{12}b_{24}}{b_{22}} x_2 \vee \frac{b_{13}b_{34}}{b_{33}} x_2 \right) = \frac{1}{b_{12}} \left( \frac{b_{12}b_{24}}{b_{22}} x_2 \vee \frac{b_{12}b_{34}}{b_{33}} x_3 \right) = \frac{b_{24}}{b_{22}} x_2 \vee \frac{b_{34}}{b_{33}} x_3.$$

Thus  $F_{(X_1, X_4) | (X_2, X_3)}(y_1, y_4 | x_2, x_3)$  for  $(x_2, x_3) \in \mathbf{X}_O(\Omega_1^o)$  coincides with the one calculated in Example 6.2.

Applying first Theorem 6.3 and then Lemma 6.5 we obtain for  $(y_1, y_4) \in \mathbb{R}_+^2$  and  $(x_2, x_3) \in (\mathbf{X}_O(\Omega_2^c) \cap \text{supp}(\mathbf{X}_O))$ ,

$$\begin{aligned} F_{(X_1, X_4)|X_2, X_3}(x_1, x_4 | x_2, x_3) &= \mathbf{1}_{[0, y_4]} \left( \frac{b_{24}}{b_{22}} x_2 \vee \frac{b_{34}}{b_{33}} x_3 \right) F_{Z_1, Z_4|X_2} \left( \frac{y_1}{b_{11}}, \frac{y_4}{b_{44}} | x_2, x_3 \right) \\ &= \mathbf{1}_{[0, y_4]} \left( \frac{b_{24}}{b_{22}} x_2 \vee \frac{b_{34}}{b_{33}} x_3 \right) F_{Z_4} \left( \frac{y_4}{b_{44}} \right) F_{Z_1|X_2} \left( \frac{y_1}{b_{11}} | x_2, x_3 \right). \end{aligned}$$

For the last term we have by Lemma 6.6,

$$F_{Z_1|X_2} \left( \frac{y_1}{b_{11}} | x_2, x_3 \right) = \frac{\frac{1}{b_{12}} \mathbf{1}_{(0, \frac{y_1}{b_{11}}]} \left( \frac{x_2}{b_{12}} \right) f_{Z_1} \left( \frac{x_2}{b_{12}} \right) F_{Z_2} \left( \frac{x_2}{b_{22}} \right) + \frac{1}{b_{22}} f_{Z_2} \left( \frac{x_2}{b_{22}} \right) F_{Z_1} \left( \frac{x_2}{b_{12}} \wedge \frac{y_1}{b_{11}} \right)}{\frac{1}{b_{12}} f_{Z_1} \left( \frac{x_2}{b_{12}} \right) F_{Z_2} \left( \frac{x_2}{b_{22}} \right) + \frac{1}{b_{22}} f_{Z_2} \left( \frac{x_2}{b_{22}} \right) F_{Z_1} \left( \frac{x_2}{b_{12}} \right)},$$

what we have also calculated in Example 6.2.  $\square$

## Appendix A: Auxiliary lemmata

**Lemma A.1.** *Let  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  be a ML graphical model, and let  $U \subseteq V$ . For coefficients  $a(i, j, k) > 0$  for  $i, j, k \in V$  we have for all  $i \in V$ ,*

$$\bigvee_{k \in \text{pa}(i)} \bigvee_{j \in \text{an}(k)} a(i, j, k) Z_j = \bigvee_{j \in \text{an}(i)} \bigvee_{k \in \text{de}(j) \cap \text{pa}(i)} a(i, j, k) Z_j, \quad (\text{A.1})$$

$$\bigvee_{k \in \text{pa}(i)} \bigvee_{j \in \text{an}(k) \setminus \text{pa}(i)} a(i, j, k) Z_j = \bigvee_{j \in \text{an}(i) \setminus \text{pa}(i)} \bigvee_{k \in \text{de}(j) \cap \text{pa}(i)} a(i, j, k) Z_j, \quad (\text{A.2})$$

$$\bigvee_{j \in \text{pa}(i) \setminus \text{pa}^{\text{tr}}(i)} \bigvee_{k \in \text{de}(j) \cap \text{pa}(i)} a(i, j, k) Z_j = \bigvee_{k \in \text{pa}(i)} \bigvee_{j \in \text{an}(k) \cap \text{pa}(i) \setminus \text{pa}^{\text{tr}}(i)} a(i, j, k) Z_j, \quad (\text{A.3})$$

$$\bigvee_{k \in \text{an}(i) \cap U} \bigvee_{j \in \text{An}(k)} a(i, j, k) Z_j = \bigvee_{j \in \text{an}(i)} \bigvee_{k \in \text{De}(j) \cap \text{an}(i) \cap U} a(i, j, k) Z_j. \quad (\text{A.4})$$

*Proof.* Since we take maxima, we only have to prove that each combination of nodes  $(k, j)$  on the left-hand side appears also on the right-hand side and vice versa. In order to prove (A.1), it therefore suffices to show

$$k \in \text{pa}(i) \text{ and } j \in \text{an}(k) \iff j \in \text{an}(i) \text{ and } k \in \text{de}(j) \cap \text{pa}(i).$$

By observing that  $j \in \text{an}(k)$  if and only if  $k \in \text{de}(j)$  and  $\text{an}(\text{pa}(i)) \subseteq \text{an}(i)$  this equivalence is obvious. The other identities are proved in the same way.  $\square$

**Lemma A.2.** *Let  $(\mathcal{D}, \mathcal{L}(\mathbf{X}))$  be a ML graphical model, and let  $U \subseteq V$ .*

- (a) *For  $i, j \in V$  such that  $j \in \text{an}(i)$ , there exists a max-weighted path from  $j$  to  $i$  containing some node of  $U \setminus \{i\}$  if and only if there exists a max-weighted path from  $j$  to  $i$  containing some node of  $\text{an}_{\text{low}}^U(i)$ .*
- (b) *For  $i, j \in V$  such that  $i \in \text{de}(j)$ , there exists a max-weighted path from  $j$  to  $i$  containing some node of  $U \setminus \{j\}$  if and only if there exists a max-weighted path from  $j$  to  $i$  containing some node of  $\text{de}_{\text{high}}^U(j)$ .*

*Proof.* We only show (a), since we may prove (b) analogously. If there exists a max-weighted path from  $j$  to  $i$  containing some node of  $\text{an}_{\text{low}}^U(i)$ , then this path contains also a node of  $U \setminus \{i\}$ , as  $\text{an}_{\text{low}}^U(i) \subseteq U \setminus \{i\}$ .

To prove the converse we construct for  $j \in \text{an}_{\text{mw}}^{U \setminus \{i\}}(i)$  a max-weighted path from  $j$  to  $i$  containing some node of  $\text{an}_{\text{low}}^U(i)$ . Let  $p = [j = k_0 \rightarrow k_1 \rightarrow \dots \rightarrow k_n = i]$  be a max-weighted path from  $j$  to  $i$  with maximal number of nodes in  $U \setminus \{i\}$  of all max-weighted path from  $j$  to  $i$ . Denote by  $l$  the lowest node on  $p$  contained in  $U \setminus \{i\}$ ; i.e., for some  $s \in \{0, 1, \dots, n\}$ , the sub-path  $p_1 = [l = k_s \rightarrow k_{s+1} \rightarrow \dots \rightarrow k_n = i]$  contains no node of  $U \setminus \{i\}$ . Assume that  $l \notin \text{an}_{\text{low}}^U(i)$ . Since  $l \in U \setminus \text{an}_{\text{low}}^U(i)$ , there exists a max-weighted path  $p_2$  from  $l$  to  $i$  containing some node in  $U \setminus \{l, i\}$ . Thus by replacing in  $p$  the sub-path  $p_1$  by  $p_2$  we obtain by Remark 2.7(iv) a max-weighted path from  $j$  to  $i$  containing more nodes of  $U \setminus \{i\}$  than  $p$ . This is, however, a contradiction to the fact that  $p$  has maximal number of nodes in  $U \setminus \{i\}$  of all max-weighted path from  $j$  to  $i$ . Hence,  $l \in \text{an}_{\text{low}}^U(i)$ , and  $p$  is a max-weighted path from  $j$  to  $i$  containing some node of  $\text{an}_{\text{low}}^U(i)$ .  $\square$

**Lemma A.3.** Assume the same situation as in Theorem 6.3. Set  $E := A \setminus \overline{O}$ . Then for  $\mathbf{y}_E \in \mathbb{R}_+^{|E|}$  we have

$$\bigvee_{\substack{t=1, \dots, T: \\ (\text{an}_{\text{mw}}^{\overline{O}_t}(i) \cap \text{cAn}^o(O_t)) \neq \emptyset}} \frac{b_{j_{t,i}^{\text{mw}}}}{b_{j_{t,i}^{\text{mw}}}} X_{i_t} \vee \bigvee_{j \in \text{An}_{\text{nmw}}^{\overline{O}}(i)} b_{ji} Z_j \leq y_i \quad \text{for all } i \in E \quad (\text{A.5})$$

if and only if

$$X_{i_t} \leq \bigwedge_{i \in \text{de}_{\text{mw}}^{\overline{O}_t}(\text{cAn}^o(O_t)) \cap E} \frac{b_{j_{t,i}^{\text{mw}}}}{b_{j_{t,i}^{\text{mw}}}} y_i \quad \text{for all } t \in \{t \in \{1, \dots, T\} : \text{an}_{\text{mw}}^{\overline{O}_t}(E) \cap \text{cAn}^o(O_t) \neq \emptyset\}$$

and

$$Z_j \leq \bigwedge_{i \in \text{De}_{\text{nmw}}^{\overline{O}}(j) \cap E} \frac{y_i}{b_{ji}} \quad \text{for all } j \in \text{An}_{\text{nmw}}^{\overline{O}}(E).$$

*Proof.* The inequalities in (A.5) are equivalent to

$$X_{i_t} \leq \frac{b_{j_{t,i}^{\text{mw}}}}{b_{j_{t,i}^{\text{mw}}}} y_i \quad \text{for all } i \in E \text{ and } t \in \{t \in \{1, \dots, T\} : \text{an}_{\text{mw}}^{\overline{O}_t}(i) \cap \text{cAn}^o(O_t) \neq \emptyset\}$$

and

$$Z_j \leq \frac{y_i}{b_{ji}} \quad \text{for all } i \in E \text{ and } j \in \text{An}_{\text{nmw}}^{\overline{O}}(i),$$

We have to show that

$$i \in E \text{ and } t \in \{t \in \{1, \dots, T\} : \text{an}_{\text{mw}}^{\overline{O}_t}(i) \cap \text{cAn}^o(O_t) \neq \emptyset\}$$

if and only if

$$t \in \{t \in \{1, \dots, T\} : \text{an}_{\text{mw}}^{\overline{O}_t}(E) \cap \text{cAn}^o(O_t) \neq \emptyset\} \text{ and } i \in \text{de}_{\text{mw}}^{\overline{O}_t}(\text{cAn}^o(O_t)) \cap E.$$

This holds true by observing that  $\text{an}_{\text{mw}}^{\overline{O}_t}(i) \cap \text{cAn}^o(O_t) \neq \emptyset$  if and only if  $i \in \text{de}_{\text{mw}}^{\overline{O}_t}(\text{cAn}^o(O_t))$ . Similarly, we obtain that

$$i \in E \text{ and } j \in \text{An}_{\text{nmw}}^{\overline{O}}(i) \quad \text{if and only if} \quad j \in \text{An}_{\text{nmw}}^{\overline{O}}(E) \text{ and } i \in \text{De}_{\text{nmw}}^{\overline{O}}(j) \cap E.$$

From these equivalences and the inequalities above assertion holds.  $\square$

## References

- [1] A. V. Aho, M. R. Garey, and J. D. Ullman. The transitive reduction of a directed graph. *SIAM Journal on Computing*, 1(2):131–137, 1972.
- [2] L. De Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer, New York, 2006.
- [3] C. Dombry, M. Ribatet, and S. Stoev. Probabilities of concurrent events. 2015. URL <http://arxiv.org/pdf/1503.05748.pdf>.
- [4] A. Klenke. *Probability Theory: A Comprehensive Course*. Springer, 2007.
- [5] D. Koller and N. Friedman. *Probabilistic Graphical Models*. MIT Press, 2009.
- [6] S. L. Lauritzen. *Graphical Models*. Oxford University Press, Oxford, 1996.
- [7] D. M. Moyles and G. L. Thompson. An algorithm for finding a minimum equivalent graph of a digraph. *Journal of the ACM*, 16(3):455–460, 1969.
- [8] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, 2nd edition, 2009.
- [9] J. Peters and P. Bühlmann. Identifiability of Gaussian structural equation models with equal error variances. *Biometrika*, 2013.
- [10] J. Peters, J. Mooij, D. Janzing, and B. Schölkopf. Causal discovery with continuous additive noise models. *Journal of Machine Learning Research*, 15:2009–2053, 2014.
- [11] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search*. MIT Press, New York, 2 edition, 2000.
- [12] Y. Wang and S. A. Stoev. Conditional sampling for spectrally discrete max-stable random fields. *Advances in Applied Probability*, 43(2):461–483, 2011.